PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7
C12N 15/00

A2

(11) International Publication Number: WO 00/65040

(43) International Publication Date: 2 November 2000 (02.11.00)

(21) International Application Number: PCT/US00/09975

(22) International Filing Date: 13 April 2000 (13.04.00)

(30) Priority Data:
60/130,849 22 April 1999 (22.04.99) US

(71) Applicant: PIONEER HI-BRED INTERNATIONAL, INC. [US/US]; 800 Capital Square, 400 Locust Street, Des Moines, IA 50309 (US).

(72) Inventors: HELENTJARIS, Timothy, G.; 2960 N.W. 73rd Lane, Ankeny, IA 50021 (US). HABBEN, Jeffrey, E.; 8813 Sunflower Circle, Urbandale, IA 50322 (US). SUN, Yuejin; 5768 N.W. 92nd Street, Johnston, IA 50131 (US).

(74) Agents: VARLEY, Karen, K. et al.; 7100 N.W. 62nd Avenue, Darwin Building, Johnston, IA 50131-1000 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

Without international search report and to be republished upon receipt of that report.

With an indication in relation to deposited biological material furnished under Rule 13bis separately from the description.

(54) Title: CELL CYCLE GENES AND METHODS OF USE

(57) Abstract

The invention provides isolated nucleic acids and their encoded proteins which are involved in cell cycle regulation. The present invention provides methods and compositions relating to altering cyclin and/or cyclin-dependent kinase concentration and/or composition of plants. The invention further provides recombinant expression cassettes, host cells, and transgenic plants.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

CELL CYCLE GENES AND METHODS OF USE

TECHNICAL FIELD

The present invention relates generally to plant molecular biology. More specifically, it relates to nucleic acids and methods for modulating their expression in plants.

BACKGROUND INFORMATION

Cell division plays a crucial role during all phases of plant development. The continuation of organogenesis and growth responses to a changing environment require precise spatial, temporal and developmental regulation of cell division activity in meristems (and in cells with the capability to form new meristems, such as in lateral root formation). Control of cell division is also important in organs themselves (i.e. separate from meristems *per se*), for example, in leaf expansion, secondary growth, and endoreduplication.

A complex network controls cell division in eukaryotes. Various regulatory pathways communicate environmental constraints such as nutrient availability, mitogenic signals such as growth factors or hormones, or developmental cues such as the transition from vegetative to reproductive growth. Ultimately, these regulatory pathways control the timing, rate, plane and position of cell division.

A plant somatic cell cycle consists of four phases: G1, a time of cell growth before DNA replication; S, a period during which DNA is replicated; G2, a period after DNA replication during which the cell prepares for division; and M, mitosis. Specialized plant cells undergo meiosis, followed by one or more cycles of mitosis, to form haploid gametophytes. Thus, tissue-preferred expression of cell cycle genes could lead to such improvements as enhanced understanding and control of transformation efficiencies, plant development, and plant fertility.

Cell division in higher eukaryotes is controlled by two main checkpoints in the cell cycle that prevent the cell from entering either M- or S-phase prematurely. Evidence from yeast and mammalian systems has repeatedly shown that over-expression of key cell cycle genes can either trigger cell division in non-dividing cells, or stimulate division in previously dividing cells (i.e., the duration of the cell cycle is decreased and cell size is reduced). Stimulation of cell division has been shown to result from over-expression of certain genes, including cyclins (see e.g. Doerner, P. et al., *Nature* (1996) 380:520-523;

Wang, T.C., et al., *Nature* (1994) 369:669-671; Quelle, D.E., *et al.*, *Genes Dev.* (1993) 7:1559-1571); E2F transcription factors (see, e.g. Johnson, D.G. *et al.*, *Nature* (1993) 365:349-352; Lukas, J. *et al.*, *Mol. Cell. Biol.* (1996) 16:1047-1057) , cdc25 (see e.g. Bell, M.H. *et al.*, *Plant Mol. Bio.* (1993) 23:445-451; Draetta, D. *et al.*, *BBA* (1996) 1332:53-63), and mdm2 (see, e.g. Teoh, G. *et al.*, *Blood* (1997) 90:1982-1992). Conversely, other gene products have been found to participate in checkpoint control, effectively blocking or retarding progression through the cell cycle (see MacLachlan, T.K. *et al.*, *Critical Rev. Eukaryotic Gene Expression* (1995) 5(2):127-156).

The basic mechanism of cell cycle control is conserved among eukaryotes. A catalytic protein kinase and an activating cyclin subunit control progress through the cell cycle. The protein kinase is generally referred to as a cyclin-dependent-kinase (CDK); its activity is modulated by phosphorylation and dephosphorylation events and by association with regulatory subunits called cyclins. CDKs require association with cyclins for activation, and the timing of activation is largely dependent upon cyclin expression.

Eukaryote genomes typically encode multiple cyclin and CDK genes. In higher eukaryotes, different members of the CDK family act in different stages of the cell cycle. Cyclin genes are classified according to the timing of their appearance during the cell cycle. In addition to cyclin and CDK subunits, CDKs are often physically associated with other proteins which alter localization, substrate specificity, or activity. A few examples of such CDK interacting proteins are the CDK inhibitors, members of the Retinoblastoma-associated protein (Rb) family, and the Constitutive Kinase Subunit (CKS).

The protein kinase activity of the complex is regulated by feedback control at certain checkpoints. At such checkpoints the CDK activity becomes limiting for further progress. When the feedback control network senses the completion of a checkpoint, CDK is activated and the cell passes through to the next checkpoint. Changes in CDK activity are regulated at multiple levels, including reversible phosphorylation of the cell cycle factors, changes in subcellular localization of the complex, and the rates of synthesis and destruction of limiting components. Regulation of the cell cycle by the cyclin/CDK complex is noted particularly at the G1/S phase transition and at the G2/M phase transition. P.W. Doerner, *Cell Cycle Regulation in Plants*, *Plant Physiol.* (1994) 106:823-827.

Plants have unique developmental features that distinguish them from other eukaryotes. Plant cells do not migrate, and thus only cell division, expansion and programmed cell death determine morphogenesis. Organs are formed throughout the

entire life span of the plant from specialized regions called meristems. In addition, many differentiated cells have the potential both to dedifferentiate and to reenter the cell cycle. There are also numerous examples of plant cell types that undergo endoreduplication, a process involving nuclear multiplication without cytokinesis. The study of plant cell cycle control genes is expected to contribute to the understanding of these unique phenomena. O. Shaul *et al.*, *Regulation of Cell Division in Arabidopsis*, *Critical Reviews in Plant Sciences*, 15(2):97-112 (1996).

Current methods for genetic engineering in maize require a specific cell type as the recipient of new DNA. These cells are found in relatively undifferentiated, rapidly growing callus cells or on the scutellar surface of the immature embryo (which gives rise to callus). There is evidence to suggest that cells must be dividing for transformation to occur. Therefore, to optimize transformation it would be desirable to provide a method for increasing the number of cells undergoing division.

It has also been observed that dividing cells represent only a fraction of cells that transiently express a transgene. Regardless of the delivery method currently used, DNA is introduced into literally thousands of cells, yet transformants are recovered at frequencies of 10^{-5} relative to transiently-expressing cells. The presence of damaged DNA in non-plant systems (similar to DNA introduced by particle gun or other physical means) has been well documented to rapidly induce cell cycle arrest. Siede, W. *Cell cycle arrest in response to DNA damage: lessons from yeast. Mutation Res.* 337(2):73-84 (1995). An increase in understanding and control of the cell cycle could help increase the rate of recovery of transformants.

Co-pending application 09/398,858 relates to CyclinD in maize, a part of the family of cyclins and cyclin-dependent kinases regulating progression through the cell cycle. Co-pending application 09/316,914 relates to cell cycle genes and in particular the constitutive kinase subunit, which interacts with cyclin-dependent kinases. Co-pending application 09/470,526 relates to genes encoding a kinase which functions at the G2/M phase transition. Identification and understanding of the polynucleotides of the instant application will aid in further understanding and control of the cell cycle.

SUMMARY OF THE INVENTION

It is an object of the present invention to provide: 1) nucleic acids and proteins relating to cell cycle genes; 2) transgenic plants comprising the nucleic acids of the present

invention; 3) methods for modulating, in a transgenic plant, the expression of the nucleic acids of the present invention.

Therefore, in one aspect, the present invention relates to an isolated nucleic acid comprising a member selected from the group consisting of (a) a polynucleotide having a specified sequence identity to a polynucleotide of the present invention; (b) a polynucleotide encoding a polypeptide of the present invention; and, (c) a polynucleotide comprising a specified number of contiguous nucleotides from a polynucleotide of (a) or (b). The isolated nucleic acid can be DNA or RNA.

In another aspect, the present invention relates to recombinant expression cassettes, comprising a nucleic acid of the present invention operably linked to a promoter.

In another aspect, the present invention is directed to a host cell into which has been introduced the recombinant expression cassette.

In a further aspect, the present invention relates to an isolated protein comprising a polypeptide having a specified number of contiguous amino acids encoded by an isolated nucleic acid of the present invention.

In a further aspect, the present invention relates to a polynucleotide amplified from a *Zea mays* nucleic acid library using primers which selectively hybridize, under stringent hybridization conditions, to loci within polynucleotides of the present invention.

In another aspect, the present invention relates to an isolated nucleic acid comprising a polynucleotide of specified length which selectively hybridizes under stringent conditions to a polynucleotide of the present invention, or a complement thereof. In some embodiments, the isolated nucleic acid is operably linked to a promoter.

In another aspect, the present invention relates to a recombinant expression cassette comprising a nucleic acid, wherein the nucleic acid is operably linked to a promoter. In some embodiments, the present invention relates to a host cell transfected with this recombinant expression cassette. In some embodiments, the present invention relates to a protein of the present invention that is produced from this host cell.

In yet another aspect, the present invention relates to a transgenic plant compr[illegible] a recombinant expression cassette comprising a promoter functional in plants operably linked to any of the isolated nucleic acids of the present invention. The present invention also provides transgenic seed from the transgenic plant.

In yet another aspect, the present invention relates to a method of modulating the level of cell cycle gene activity in a plant cell capable of plant regeneration.

Definitions

Units, prefixes, and symbols may be denoted in their SI accepted form. Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation; amino acid sequences are written left to right in amino to carboxy orientation, respectively. Numeric ranges are inclusive of the numbers defining the range and include each integer within the defined range. Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes. Unless otherwise provided for, software, electrical, and electronics terms as used herein are as defined in The New IEEE Standard Dictionary of Electrical and Electronics Terms (5th edition, 1993). The terms defined below are more fully defined by reference to the specification as a whole.

By "amplified" is meant the construction of multiple copies of a nucleic acid sequence or multiple copies complementary to the nucleic acid sequence using at least one of the nucleic acid sequences as a template. Amplification systems include the polymerase chain reaction (PCR) system, ligase chain reaction (LCR) system, nucleic acid sequence based amplification (NASBA, Cangene, Mississauga, Ontario), Q-*Beta* Replicase systems, transcription-based amplification system (TAS), and strand displacement amplification (SDA). See, e.g., *Diagnostic Molecular Microbiology: Principles and Applications*, D. H. Persing *et al.*, Ed., American Society for Microbiology, Washington, D.C. (1993). The product of amplification is termed an amplicon.

The term "antibody" includes reference to antigen binding forms of antibodies (e.g., Fab, $F(ab)_2$). The term "antibody" frequently refers to a polypeptide substantially encoded by an immunoglobulin gene or immunoglobulin genes, or fragments thereof which specifically bind and recognize an analyte (antigen). However, while various antibody fragments can be defined in terms of the digestion of an intact antibody, one of skill will appreciate that such fragments may be synthesized *de novo* either chemically or by utilizing recombinant DNA methodology. Thus, the term antibody, as used herein, also includes antibody fragments such as single chain Fv, chimeric antibodies (i.e., comprising constant and variable regions from different species), humanized antibodies (i.e., comprising a complementarity determining region (CDR) from a non-human source) and heteroconjugate antibodies (e.g., bispecific antibodies).

As used herein, "antisense orientation" includes reference to a duplex polynucleotide sequence that is operably linked to a promoter in an orientation where the antisense strand is transcribed. The antisense strand is sufficiently complementary to an endogenous transcription product such that translation of the endogenous transcription product is often inhibited.

As used herein, "chromosomal region" includes reference to a length of a chromosome that may be measured by reference to the linear segment of DNA that it comprises. The chromosomal region can be defined by reference to two unique DNA sequences, i.e., markers.

The term "conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refers to those nucleic acids which encode identical or conservatively modified variants of the amino acid sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations" and represent one species of conservatively modified variation. Every nucleic acid sequence herein that encodes a polypeptide also, by reference to the genetic code, describes every possible silent variation of the nucleic acid. One of ordinary skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine; and UGG , which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid which encodes a polypeptide of the present invention is implicit in each described polypeptide sequence and is within the scope of the present invention.

As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Thus, any number of amino acid residues selected from the group of integers consisting of from 1 to 15 can be so altered. Thus, for example, 1, 2, 3, 4, 5, 7, or 10 alterations can be made.

Conservatively modified variants typically provide similar biological activity as the unmodified polypeptide sequence from which they are derived. For example, substrate specificity, enzyme activity, or ligand/receptor binding is generally at least 30%, 40%, 50%, 60%, 70%, 80%, or 90% of the native protein for its native substrate. Conservative substitution tables providing functionally similar amino acids are well known in the art.

The following six groups each contain amino acids that are conservative substitutions for one another:

1) Alanine (A), Serine (S), Threonine (T);
2) Aspartic acid (D), Glutamic acid (E);
3) Asparagine (N), Glutamine (Q);
4) Arginine (R), Lysine (K);
5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and
6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

See also, Creighton (1984) Proteins W.H. Freeman and Company.

By "encoding" or "encoded", with respect to a specified nucleic acid, is meant comprising the information for translation into the specified protein. A nucleic acid encoding a protein may comprise non-translated sequences (e.g., introns) within translated regions of the nucleic acid, or may lack such intervening non-translated sequences (e.g., as in cDNA). The information by which a protein is encoded is specified by the use of codons. Typically, the amino acid sequence is encoded by the nucleic acid using the "universal" genetic code. However, variants of the universal code, such as are present in some plant, animal, and fungal mitochondria, the bacterium *Mycoplasma capricolum*, or the ciliate *Macronucleus*, may be used when the nucleic acid is expressed therein.

When the nucleic acid is prepared or altered synthetically, advantage can be taken of known codon preferences of the intended host where the nucleic acid is to be expressed. For example, although nucleic acid sequences of the present invention may be expressed in both monocotyledonous and dicotyledonous plant species, sequences can be modified to account for the specific codon preferences and GC content preferences of monocotyledons or dicotyledons as these preferences have been shown to differ (Murray *et al. Nucl. Acids Res.* 17: 477-498 (1989)). Thus, the maize preferred codon for a particular amino acid may be derived from known gene sequences from maize. Maize codon usage for 28 genes from maize plants are listed in Table 4 of Murray *et al.*, *supra*.

As used herein "full-length sequence" in reference to a specified polynucleotide or its encoded protein means having the entire amino acid sequence of, a native (non-synthetic), endogenous, biologically active form of the specified protein. Methods to determine whether a sequence is full-length are well known in the art including such exemplary techniques as northern or western blots, primer extension, S1 protection, and ribonuclease protection. See, e.g., *Plant Molecular Biology: A Laboratory Manual*, Clark, Ed., Springer-Verlag, Berlin (1997). Comparison to known full-length homologous (orthologous and/or paralogous) sequences can also be used to identify full-length sequences of the present invention. Additionally, consensus sequences typically present at the 5' and 3' untranslated regions of mRNA aid in the identification of a polynucleotide as full-length. For example, the consensus sequence ANNNNAUGG, where the underlined codon represents the N-terminal methionine, aids in determining whether the polynucleotide has a complete 5' end. Consensus sequences at the 3' end, such as polyadenylation sequences, aid in determining whether the polynucleotide has a complete 3' end.

The term "gene activity" refers to one or more steps involved in gene expression, including transcription, translation, and the functioning of the protein encoded by the gene.

As used herein, "heterologous" in reference to a nucleic acid is a nucleic acid that originates from a foreign species, or, if from the same species, is substantially modified from its native form in composition and/or genomic locus by deliberate human intervention. For example, a promoter operably linked to a heterologous structural gene is from a species different from that from which the structural gene was derived, or, if from the same species, one or both are substantially modified from their original form. A heterologous protein may originate from a foreign species or, if from the same species, is substantially modified from its original form by deliberate human intervention.

By "host cell" is meant a cell which contains a vector and supports the replication and/or expression of the vector. Host cells may be prokaryotic cells such as *E. coli*, or eukaryotic cells such as yeast, insect, amphibian, or mammalian cells. Preferably, host cells are monocotyledonous or dicotyledonous plant cells. A particularly preferred monocotyledonous host cell is a maize host cell.

The term "hybridization complex" includes reference to a duplex nucleic acid structure formed by two single-stranded nucleic acid sequences selectively hybridized with each other.

The term "introduced" in the context of inserting a nucleic acid into a cell, means "transfection" or "transformation" or "transduction" and includes reference to the incorporation of a nucleic acid into a eukaryotic or prokaryotic cell where the nucleic acid may be incorporated into the genome of the cell (e.g., chromosome, plasmid, plastid or mitochondrial DNA), converted into an autonomous replicon, or transiently expressed (e.g., transfected mRNA).

The term "isolated" refers to material, such as a nucleic acid or a protein, which is: (1) substantially or essentially free from components that normally accompany or interact with it as found in its naturally occurring environment. The isolated material optionally comprises material not found with the material in its natural environment; or (2) if the material is in its natural environment, the material has been synthetically (non-naturally) altered by deliberate human intervention to a composition and/or placed at a location in the cell (e.g., genome or subcellular organelle) not native to a material found in that environment. The alteration to yield the synthetic material can be performed on the material within or removed from its natural state. For example, a naturally occurring nucleic acid becomes an isolated nucleic acid if it is altered, or if it is transcribed from DNA which has been altered, by means of human intervention performed within the cell from which it originates. See, e.g., *Compounds and Methods for Site Directed Mutagenesis in Eukaryotic Cells*, Kmiec, U.S. Patent No. 5,565,350; *In Vivo Homologous Sequence Targeting in Eukaryotic Cells*; Zarling *et al.*, WO 93/22443. Likewise, a naturally occurring nucleic acid (e.g., a promoter) becomes isolated if it is introduced by non-naturally occurring means to a locus of the genome not native to that nucleic acid. Nucleic acids which are "isolated" as defined herein, are also referred to as "heterologous" nucleic acids.

Unless otherwise stated, the term "cell cycle nucleic acid" refers to a nucleic acid of the present invention and means a nucleic acid comprising a polynucleotide of the present invention (a "cell cycle polynucleotide") encoding a cell cycle polypeptide. A "cell cycle gene" is a gene of the present invention and refers to a full-length cell cycle polynucleotide.

As used herein, "localized within the chromosomal region defined by and including" with respect to particular markers includes reference to a contiguous length of a chromosome delimited by and including the stated markers.

As used herein, "marker" includes reference to a locus on a chromosome that serves to identify a unique position on the chromosome. A "polymorphic marker" includes reference to a marker which appears in multiple forms (alleles) such that different forms of the marker, when they are present in a homologous pair, allow transmission of each of the chromosomes of that pair to be followed. A genotype may be defined by use of one or a plurality of markers.

As used herein, "nucleic acid" includes reference to a deoxyribonucleotide or ribonucleotide polymer in either single- or double-stranded form, and unless otherwise limited, encompasses known analogues having the essential nature of natural nucleotides in that they hybridize to single-stranded nucleic acids in a manner similar to naturally occurring nucleotides (e.g., peptide nucleic acids).

By "nucleic acid library" is meant a collection of isolated DNA or RNA molecules which comprise and substantially represent the entire transcribed fraction of a genome of a specified organism. Construction of exemplary nucleic acid libraries, such as genomic and cDNA libraries, is taught in standard molecular biology references such as Berger and Kimmel, *Guide to Molecular Cloning Techniques, Methods in Enzymology*, Vol. 152, Academic Press, Inc., San Diego, CA, (Berger); Sambrook *et al., Molecular Cloning - A Laboratory Manual,* 2nd ed., Vol. 1-3 (1989); and *Current Protocols in Molecular Biology*, F.M. Ausubel *et al.*, Eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc. (1994).

As used herein "operably linked" includes reference to a functional linkage between a promoter and a second sequence, wherein the promoter sequence initiates and mediates transcription of the DNA sequence corresponding to the second sequence. Generally, operably linked means that the nucleic acid sequences being linked are contiguous and, where necessary to join two protein coding regions, contiguous and in the same reading frame.

As used herein, the term "plant" can include reference to whole plants, plant parts or organs (e.g., leaves, stems, roots, etc.), plant cells, seeds and progeny of same. Plant cell, as used herein, further includes, without limitation, cells obtained from or found in: seeds, suspension cultures, embryos, meristematic regions, callus tissue, leaves, roots, shoots, gametophytes, sporophytes, pollen, and microspores. Plant cells can also be understood to include modified cells, such as protoplasts, obtained from the aforementioned tissues. The class of plants which can be used in the methods of the

invention is generally as broad as the class of higher plants amenable to transformation techniques, including both monocotyledonous and dicotyledonous plants. Particularly preferred plants include maize, soybean, sunflower, sorghum, canola, wheat, alfalfa, cotton, rice, barley, and millet.

As used herein, "polynucleotide" includes reference to a deoxyribopolynucleotide, ribopolynucleotide, or analogs thereof that have the essential nature of a natural ribonucleotide in that they hybridize, under stringent hybridization conditions, to substantially the same nucleotide sequence as naturally occurring nucleotides and/or allow translation into the same amino acid(s) as the naturally occurring nucleotide(s). A polynucleotide can be full-length or a subsequence of a native or heterologous structural or regulatory gene. Unless otherwise indicated, the term includes reference to the specified sequence as well as the complementary sequence thereof. Thus, DNAs or RNAs with backbones modified for stability or for other reasons are "polynucleotides" as that term is intended herein. Moreover, DNAs or RNAs comprising unusual bases, such as inosine, or modified bases, such as tritylated bases, to name just two examples, are polynucleotides as the term is used herein. It will be appreciated that a great variety of modifications have been made to DNA and RNA that serve many useful purposes known to those of skill in the art. The term polynucleotide as it is employed herein embraces such chemically, enzymatically or metabolically modified forms of polynucleotides, as well as the chemical forms of DNA and RNA characteristic of viruses and cells, including among other things, simple and complex cells.

The terms "polypeptide", "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical analogue of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers. The essential nature of such analogues of naturally occurring amino acids is that, when incorporated into a protein, that protein is specifically reactive to antibodies elicited to the same protein but consisting entirely of naturally occurring amino acids. The terms "polypeptide", "peptide" and "protein" are also inclusive of modifications including, but not limited to, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation. It will be appreciated, as is well known and as noted above, that polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of ubiquitination, and they may be circular, with or without

branching, generally as a result of posttranslation events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translation natural process and by entirely synthetic methods, as well. Further, this invention contemplates the use of both the methionine-containing and the methionine-less amino terminal variants of the protein of the invention.

As used herein "promoter" includes reference to a region of DNA upstream from the start of transcription and involved in recognition and binding of RNA polymerase and other proteins to initiate transcription. A "plant promoter" is a promoter capable of initiating transcription in plant cells whether nor not its origin is a plant cell. Exemplary plant promoters include, but are not limited to, those that are obtained from plants, plant viruses, and bacteria which comprise genes expressed in plant cells such *Agrobacterium* or *Rhizobium*. Examples of promoters under developmental control include promoters that preferentially initiate transcription in certain tissues, such as leaves, roots, or seeds. Such promoters are referred to as "tissue preferred". Promoters which initiate transcription only in certain tissue are referred to as "tissue specific". A "cell type" specific promoter primarily drives expression in certain cell types in one or more organs, for example, vascular cells in roots or leaves. An "inducible" or "repressible" promoter is a promoter which is under environmental control. Examples of environmental conditions that may effect transcription by inducible promoters include anaerobic conditions or the presence of light. Tissue specific, tissue preferred, cell type specific, and inducible promoters constitute the class of "non-constitutive" promoters. A "constitutive" promoter is a promoter which is active under most environmental conditions.

The term "cell cycle polypeptide" is a polypeptide of the present invention and refers to one or more amino acid sequences, in glycosylated or non-glycosylated form. The term is also inclusive of fragments, variants, homologs, alleles or precursors (e.g., preproproteins or proproteins) thereof. A "cell cycle protein" is a protein of the present invention and comprises a cell cycle polypeptide.

As used herein "recombinant" includes reference to a cell or vector, that has been modified by the introduction of a heterologous nucleic acid or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found in identical form within the native (non-recombinant) form of the cell or express native genes that are otherwise abnormally expressed, under-expressed or not expressed at all as a

result of deliberate human intervention. The term "recombinant" as used herein does not encompass the alteration of the cell or vector by naturally occurring events (e.g., spontaneous mutation, natural transformation/transduction/transposition) such as those occurring without deliberate human intervention.

As used herein, a "recombinant expression cassette" is a nucleic acid construct, generated recombinantly or synthetically, with a series of specified nucleic acid elements which permit transcription of a particular nucleic acid in a host cell. The recombinant expression cassette can be incorporated into a plasmid, chromosome, mitochondrial DNA, plastid DNA, virus, or nucleic acid fragment. Typically, the recombinant expression cassette portion of an expression vector includes, among other sequences, a nucleic acid to be transcribed, and a promoter.

The term "residue" or "amino acid residue" or "amino acid" are used interchangeably herein to refer to an amino acid that is incorporated into a protein, polypeptide, or peptide (collectively "protein"). The amino acid may be a naturally occurring amino acid and, unless otherwise limited, may encompass non-natural analogs of natural amino acids that can function in a similar manner as naturally occurring amino acids.

The term "selectively hybridizes" includes reference to hybridization, under stringent hybridization conditions, of a nucleic acid sequence to a specified nucleic acid target sequence to a detectably greater degree (e.g., at least 2-fold over background) than its hybridization to non-target nucleic acid sequences and to the substantial exclusion of non-target nucleic acids. Selectively hybridizing sequences typically have about at least 80% sequence identity, preferably 90% sequence identity, and most preferably 100% sequence identity (i.e., complementary) with each other.

The term "stringent conditions" or "stringent hybridization conditions" includes reference to conditions under which a probe will hybridize to its target sequence, to a detectably greater degree than to other sequences (e.g., at least 2-fold over background). Stringent conditions are sequence-dependent and will be different in different circumstances. By controlling the stringency of the hybridization and/or washing conditions, target sequences can be identified which are 100% complementary to the probe (homologous probing). Alternatively, stringency conditions can be adjusted to allow some mismatching in sequences so that lower degrees of similarity are detected (heterologous

probing). Generally, a probe is less than about 1000 nucleotides in length, optionally less than 500 nucleotides in length.

Typically, stringent conditions will be those in which the salt concentration is less than about 1.5 M Na ion, typically about 0.01 to 1.0 M Na ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (*e.g.*, 10 to 50 nucleotides) and at least about 60°C for long probes (*e.g.*, greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. Exemplary low stringency conditions include hybridization with a buffer solution of 30 to 35% formamide, 1 M NaCl, 1% SDS (sodium dodecyl sulphate) at 37°C, and a wash in 1X to 2X SSC (20X SSC = 3.0 M NaCl/0.3 M trisodium citrate) at 50 to 55°C. Exemplary moderate stringency conditions include hybridization in 40 to 45% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 0.5X to 1X SSC at 55 to 60°C. Exemplary high stringency conditions include hybridization in 50% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 0.1X SSC at 60 to 65°C.

Specificity is typically the function of post-hybridization washes, the critical factors being the ionic strength and temperature of the final wash solution. For DNA-DNA hybrids, the T_m can be approximated from the equation of Meinkoth and Wahl, *Anal. Biochem.*, 138:267-284 (1984): $T_m = 81.5\ °C + 16.6\ (\log M) + 0.41\ (\%GC) - 0.61\ (\%\ form) - 500/L$; where M is the molarity of monovalent cations, %GC is the percentage of guanosine and cytosine nucleotides in the DNA, % form is the percentage of formamide in the hybridization solution, and L is the length of the hybrid in base pairs. The T_m is the temperature (under defined ionic strength and pH) at which 50% of a complementary target sequence hybridizes to a perfectly matched probe. T_m is reduced by about 1 °C for each 1% of mismatching; thus, T_m, hybridization and/or wash conditions can be adjusted to hybridize to sequences of the desired identity. For example, if sequences with $\geq$90% identity are sought, the T_m can be decreased 10 °C. Generally, stringent conditions are selected to be about 5 °C lower than the thermal melting point (T_m) for the specific sequence and its complement at a defined ionic strength and pH. However, severely stringent conditions can utilize a hybridization and/or wash at 1, 2, 3, or 4 °C lower than the thermal melting point (T_m); moderately stringent conditions can utilize a hybridization and/or wash at 6, 7, 8, 9, or 10 °C lower than the thermal melting point (T_m); low stringency conditions can utilize a hybridization and/or wash at 11, 12, 13, 14, 15, or 20 °C lower than the thermal melting point (T_m). Using the equation, hybridization and wash

compositions, and desired T_m, those of ordinary skill will understand that variations in the stringency of hybridization and/or wash solutions are inherently described. If the desired degree of mismatching results in a T_m of less than 45 °C (aqueous solution) or 32 °C (formamide solution) it is preferred to increase the SSC concentration so that a higher temperature can be used. Hybridization and/or wash conditions can be applied for at least 10, 30, 60, 90, 120, or 240 minutes. An extensive guide to the hybridization of nucleic acids is found in Tijssen, *Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes,* Part I, Chapter 2 "Overview of principles of hybridization and the strategy of nucleic acid probe assays", Elsevier, New York (1993); and *Current Protocols in Molecular Biology*, Chapter 2, Ausubel, *et al.*, Eds., Greene Publishing and Wiley-Interscience, New York (1995).

As used herein, "transgenic plant" includes reference to a plant which comprises within its genome a heterologous polynucleotide. Generally, the heterologous polynucleotide is stably integrated within the genome such that the polynucleotide is passed on to successive generations. The heterologous polynucleotide may be integrated into the genome alone or as part of a recombinant expression cassette. "Transgenic" is used herein to include any cell, cell line, callus, tissue, plant part or plant, the genotype of which has been altered by the presence of heterologous nucleic acid including those transgenics initially so altered as well as those created by sexual crosses or asexual propagation from the initial transgenic. The term "transgenic" as used herein does not encompass the alteration of the genome (chromosomal or extra-chromosomal) by conventional plant breeding methods or by naturally occurring events such as random cross-fertilization, non-recombinant viral infection, non-recombinant bacterial transformation, non-recombinant transposition, or spontaneous mutation.

As used herein, "vector" includes reference to a nucleic acid used in transfection of a host cell and into which can be inserted a polynucleotide. Vectors are often replicons. Expression vectors permit transcription of a nucleic acid inserted therein.

The following terms are used to describe the sequence relationships between a polynucleotide/polypeptide of the present invention and a reference polynucleotide/polypeptide: (a) "reference sequence", (b) "comparison window", (c) "sequence identity", and (d) "percentage of sequence identity".

(a) As used herein, "reference sequence" is a defined sequence used as a basis for sequence comparison with a polynucleotide/polypeptide of the present invention. A

reference sequence may be a subset or the entirety of a specified sequence; for example, as a segment of a full-length cDNA or gene sequence, or the complete cDNA or gene sequence.

(b) As used herein, "comparison window" includes reference to a contiguous and specified segment of a polynucleotide/polypeptide sequence, wherein the polynucleotide/polypeptide sequence may be compared to a reference sequence and wherein the portion of the polynucleotide/polypeptide sequence in the comparison window may comprise additions or deletions (i.e., gaps) compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Generally, the comparison window is at least 20 contiguous nucleotides/amino acid residues in length, and optionally can be 30, 40, 50, 100, or longer. Those of skill in the art understand that to avoid a high similarity to a reference sequence due to inclusion of gaps in the polynucleotide/polypeptide sequence, a gap penalty is typically introduced and is subtracted from the number of matches.

Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison may be conducted by the local homology algorithm of Smith and Waterman, *Adv. Appl. Math.* 2: 482 (1981); by the homology alignment algorithm of Needleman and Wunsch, *J. Mol. Biol.* 48: 443 (1970); by the search for similarity method of Pearson and Lipman, *Proc. Nat'l. Acad. Sci.* 85: 2444 (1988); by computerized implementations of these algorithms, including, but not limited to: CLUSTAL in the PC/Gene program by Intelligenetics, Mountain View, California; GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 575 Science Dr., Madison, Wisconsin, USA; the CLUSTAL program is well described by Higgins and Sharp, *Gene* 73: 237-244 (1988); Higgins and Sharp, *CABIOS* 5: 151-153 (1989); Corpet, *et al., Nucleic Acids Research* 16: 10881-90 (1988); Huang, *et al., Computer Applications in the Biosciences* 8: 155-65 (1992), and Pearson, *et al., Methods in Molecular Biology* 24: 307-331 (1994).

The BLAST family of programs which can be used for database similarity searches includes: BLASTN for nucleotide query sequences against nucleotide database sequences; BLASTX for nucleotide query sequences against protein database sequences; BLASTP for protein query sequences against protein database sequences; TBLASTN for protein query sequences against nucleotide database sequences; and TBLASTX for nucleotide query sequences against nucleotide database sequences. See, *Current Protocols in Molecular*

Biology, Chapter 19, Ausubel, *et al.*, Eds., Greene Publishing and Wiley-Interscience, New York (1995); Altschul *et al.*, *J. Mol. Biol.*, 215:403-410 (1990); and, Altschul *et al.*, *Nucleic Acids Res.* 25:3389-3402 (1997).

Software for performing BLAST analyses is publicly available, e.g., through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al.*, *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (*see* Henikoff & Henikoff (1989) *Proc. Nat'l. Acad. Sci. USA* 89:10915).

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see, e.g.*, Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance.

BLAST searches assume that proteins can be modeled as random sequences. However, many real proteins comprise regions of nonrandom sequences which may be

homopolymeric tracts, short-period repeats, or regions enriched in one or more amino acids. Such low-complexity regions may be aligned between unrelated proteins even though other regions of the protein are entirely dissimilar. A number of low-complexity filter programs can be employed to reduce such low-complexity alignments. For example, the SEG (Wooten and Federhen, *Comput. Chem.*, 17:149-163 (1993)) and XNU (Claverie and States, *Comput. Chem.*, 17:191-201 (1993)) low-complexity filters can be employed alone or in combination.

Unless otherwise stated, nucleotide and protein identity/similarity values provided herein are preferably calculated using GAP (GCG Version 10) and/or BLAST under default values.

GAP (Global Alignment Program) can also be used to compare a polynucleotide or polypeptide of the present invention with a reference sequence. GAP uses the algorithm of Needleman and Wunsch (*J. Mol. Biol.* 48: 443-453, 1970) to find the alignment of two complete sequences that maximizes the number of matches and minimizes the number of gaps. GAP considers all possible alignments and gap positions and creates the alignment with the largest number of matched bases and the fewest gaps. It allows for the provision of a gap creation penalty and a gap extension penalty in units of matched bases. GAP must make a profit of gap creation penalty number of matches for each gap it inserts. If a gap extension penalty greater than zero is chosen, GAP must, in addition, make a profit for each gap inserted of the length of the gap times the gap extension penalty. Default gap creation penalty values and gap extension penalty values in Version 10 of the Wisconsin Genetics Software Package for protein sequences are 8 and 2, respectively. For nucleotide sequences the default gap creation penalty is 50 while the default gap extension penalty is 3. The gap creation and gap extension penalties can be expressed as an integer selected from the group of integers consisting of from 0 to 100. Thus, for example, the gap creation and gap extension penalties can each independently be: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60 or greater.

GAP presents one member of the family of best alignments. There may be many members of this family, but no other member has a better quality. GAP displays four figures of merit for alignments: Quality, Ratio, Identity, and Similarity. The Quality is the metric maximized in order to align the sequences. Ratio is the quality divided by the number of bases in the shorter segment. Percent Identity is the percent of the symbols that actually match. Percent Similarity is the percent of the symbols that are similar. Symbols

that are across from gaps are ignored. A similarity is scored when the scoring matrix value for a pair of symbols is greater than or equal to 0.50, the similarity threshold. The scoring matrix used in Version 10 of the Wisconsin Genetics Software Package is BLOSUM62 (*see* Henikoff & Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89:10915).

Multiple alignment of the sequences can be performed using the CLUSTAL method of alignment (Higgins and Sharp (1989) *CABIOS. 5*:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the CLUSTAL method are KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5.

(c) As used herein, "sequence identity" or "identity" in the context of two nucleic acid or polypeptide sequences includes reference to the residues in the two sequences which are the same when aligned for maximum correspondence over a specified comparison window. When percentage of sequence identity is used in reference to proteins it is recognized that residue positions which are not identical often differ by conservative amino acid substitutions, where amino acid residues are substituted for other amino acid residues with similar chemical properties (e.g. charge or hydrophobicity) and therefore do not change the functional properties of the molecule. Where sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Sequences which differ by such conservative substitutions are said to have "sequence similarity" or "similarity". Means for making this adjustment are well-known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated, *e.g.*, according to the algorithm of Meyers and Miller, *Computer Applic. Biol. Sci.*, 4: 11-17 (1988) *e.g.*, as implemented in the program PC/GENE (Intelligenetics, Mountain View, California, USA).

(d) As used herein, "percentage of sequence identity" means the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is

calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity.

DETAILED DESCRIPTION OF THE INVENTION

Overview

The present invention provides, among other things, compositions and methods for modulating (i.e., increasing or decreasing) the level of polynucleotides and polypeptides of the present invention in plants. In particular, the polynucleotides and polypeptides of the present invention can be expressed temporally or spatially, e.g., at developmental stages, in tissues, and/or in quantities, which are uncharacteristic of non-recombinantly engineered plants. Thus, the present invention provides utility in such exemplary applications as improving transformation efficiency (for example, through enhancing transgene integration and/or providing a positive growth advantage to transformed cells); altering response to pathogens, particularly those inducing cell proliferation; increasing overall crop yield (for example, through improving vigor or growth rate); and selectively modulating growth rate of specific tissues.

The present invention also provides isolated nucleic acids comprising polynucleotides of sufficient length and complementarity to a gene of the present invention to use as probes or amplification primers in the detection, quantitation, or isolation of gene transcripts. For example, isolated nucleic acids of the present invention can be used as probes in detecting deficiencies in the level of mRNA in screenings for desired transgenic plants, for detecting mutations in the gene (e.g., substitutions, deletions, or additions), for monitoring upregulation of expression or changes in enzyme activity in screening assays of compounds, for detection of any number of allelic variants (polymorphisms), orthologs, or paralogs of the gene, or for site directed mutagenesis in eukaryotic cells (see, e.g., U.S. Patent No. 5,565,350). The isolated nucleic acids of the present invention can also be used for recombinant expression of their encoded polypeptides, or for use as immunogens in the preparation and/or screening of antibodies. The isolated nucleic acids of the present invention can also be employed for use in sense or antisense suppression of one or more

genes of the present invention in a host cell, tissue, or plant. Attachment of chemical agents which bind, intercalate, cleave and/or crosslink to the isolated nucleic acids of the present invention can also be used to modulate transcription or translation.

The present invention also provides isolated proteins comprising a polypeptide of the present invention (e.g., preproenzyme, proenzyme, or enzymes). The present invention also provides proteins comprising at least one epitope from a polypeptide of the present invention. The proteins of the present invention can be employed in assays for enzyme agonists or antagonists of enzyme function, or for use as immunogens or antigens to obtain antibodies specifically immunoreactive with a protein of the present invention. Such antibodies can be used in assays for expression levels, for identifying and/or isolating nucleic acids of the present invention from expression libraries, for identification of homologous polypeptides from other species, or for purification of polypeptides of the present invention.

The isolated nucleic acids and polypeptides of the present invention can be used over a broad range of plant types, particularly monocots such as the species of the family *Gramineae* including *Hordeum, Secale, Triticum, Sorghum* (e.g., *S. bicolor*) and *Zea* (e.g., *Z. mays*). The isolated nucleic acid and proteins of the present invention can also be used in species from the genera: *Cucurbita, Rosa, Vitis, Juglans, Fragaria, Lotus, Medicago, Onobrychis, Trifolium, Trigonella, Vigna, Citrus, Linum, Geranium, Manihot, Daucus, Arabidopsis, Brassica, Raphanus, Sinapis, Atropa, Capsicum, Datura, Hyoscyamus, Lycopersicon, Nicotiana, Solanum, Petunia, Digitalis, Majorana, Ciahorium, Helianthus, Lactuca, Bromus, Asparagus, Antirrhinum, Heterocallis, Nemesis, Pelargonium, Panicum, Pennisetum, Ranunculus, Senecio, Salpiglossis, Cucumis, Browallia, Glycine, Pisum, Phaseolus, Lolium, Gossypium. Oryza,* and *Avena*.

Nucleic Acids

The present invention provides, among other things, isolated nucleic acids of RNA, DNA, and analogs and/or chimeras thereof, comprising a polynucleotide of the present invention.

A polynucleotide of the present invention is inclusive of:

(a) a polynucleotide encoding a polypeptide of SEQ ID NO: 2, 6, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, or 50, and conservatively modified and polymorphic variants

thereof, including exemplary polynucleotides of SEQ ID NOS: 1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, and 49;

(b) a polynucleotide which is the product of amplification from a *Zea mays* nucleic acid library using primer pairs which selectively hybridize under stringent conditions to loci within a polynucleotide selected from the group consisting of SEQ ID NOS: 1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, and 49, wherein the polynucleotide has substantial sequence identity to a polynucleotide selected from the group consisting of SEQ ID NOS: 1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, and 49;

(c) a polynucleotide which selectively hybridizes to a polynucleotide of (a) or (b);

(d) a polynucleotide having a specified sequence identity with polynucleotides of (a), (b), or (c);

(e) a polynucleotide encoding a protein having a specified number of contiguous amino acids from a prototype polypeptide, wherein the protein is specifically recognized by antisera elicited by presentation of the protein and wherein the protein does not detectably immunoreact to antisera which has been fully immunosorbed with the protein;

(f) complementary sequences of polynucleotides of (a), (b), (c), (d), or (e); and

(g) a polynucleotide comprising at least a specific number of contiguous nucleotides from a polynucleotide of (a), (b), (c), (d), (e), or (f).

A. Polynucleotides Encoding A Polypeptide of the Present Invention or Conservatively Modified or Polymorphic Variants Thereof

As indicated in (a), above, the present invention provides isolated nucleic acids comprising a polynucleotide of the present invention, wherein the polynucleotide encodes a polypeptide of the present invention, or conservatively modified or polymorphic variants thereof. Accordingly, the present invention includes, for example, polynucleotides of SEQ ID NOS: 1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, and 49 and silent variations of polynucleotides encoding a polypeptide of SEQ ID NO: 2, 6, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, or 50. The present invention further provides isolated nucleic acids comprising polynucleotides encoding conservatively modified variants of a polypeptide of SEQ ID NO: 2, 6, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, or 50. Additionally, the present invention further provides isolated nucleic acids comprising polynucleotides encoding one or more allelic (polymorphic) variants of polypeptides/polynucleotides. Polymorphic variants are

frequently used to follow segregation of chromosomal regions in, for example, marker assisted selection methods for crop improvement.

B. Polynucleotides Amplified from a Zea mays Nucleic Acid Library

As indicated in (b), above, the present invention provides an isolated nucleic acid comprising a polynucleotide of the present invention, wherein the polynucleotides are amplified from a *Zea mays* nucleic acid library. *Zea mays* lines B73, PHRE1, A632, BMS-P2#10, W23, and Mo17 are known and publicly available. Other publicly known and available maize lines can be obtained from the Maize Genetics Cooperation (Urbana, IL). The nucleic acid library may be a cDNA library, a genomic library, or a library generally constructed from nuclear transcripts at any stage of intron processing. cDNA libraries can be normalized to increase the representation of relatively rare cDNAs. In optional embodiments, the cDNA library is constructed using a full-length cDNA synthesis method. Examples of such methods include Oligo-Capping (Maruyama, K. and Sugano, S. *Gene* 138: 171-174, 1994), Biotinylated CAP Trapper (Carninci, P., Kvan, C., *et al. Genomics* 37: 327-336, 1996), and CAP Retention Procedure (Edery, E., Chu, L.L., *et al. Molecular and Cellular Biology* 15: 3363-3371, 1995). cDNA synthesis is often catalyzed at 50-55^{o}C to prevent formation of RNA secondary structure. Examples of reverse transcriptases that are relatively stable at these temperatures are SuperScript II Reverse Transcriptase (Life Technologies, Inc.), AMV Reverse Transcriptase (Boehringer Mannheim) and RetroAmp Reverse Transcriptase (Epicentre). Rapidly growing tissues, or rapidly dividing cells are preferably used as mRNA sources. Libraries can be made from a variety of maize tissues, but for optimal results one should isolate RNA from mitotically active tissues such as shoot meristems, shoot meristem cultures, callus and suspension cultures, immature ears and tassels, and young seedlings. Since cyclins are typically expressed at specific cell cycle stages, it may be possible to enrich for such rare messages using cell cycle inhibitors such as aphidicolin, hydroxyurea, and mimosine to block cells at the G1/S boundary. Cells can also be blocked at this stage using the double phosphate starvation method. Synchronization of source cells using intermittent periods of light and darkness may also be useful. Hormone treatments that stimulate cell division, for example cytokinin, would also increase expression of cell cycle genes.

Full length cDNA libraries from such rapidly-dividing tissues (or cells at the G1/S boundary) would provide opportunities for identifying full-length, cell-cycle-related

cDNAs. Full length cDNA libraries can be constructed using the "Biotinylated CAP Trapper" method (Carninci, P., et al, *Genomics* Vol. 37, pp. 327-336, 1996) or the "mRNA Cap Retention Procedure" (Edery, I., et al., *Molecular and Cellular Biology* Vol. 15, pp. 3363-3371, 1995). Full length cDNA libraries can be normalized to provide a higher probability of sampling genes that express at low levels. Examples of cDNA library normalization methods are summarized by Bento Soares (Bonaldo, M.F., et al., *Genome Research*, Vol. 6, pp. 791-806, 1996).

The present invention also provides subsequences of the polynucleotides of the present invention. A variety of subsequences can be obtained using primers which selectively hybridize under stringent conditions to at least two sites within a polynucleotide of the present invention, or to two sites within the nucleic acid which flank and comprise a polynucleotide of the present invention, or to a site within a polynucleotide of the present invention and a site within the nucleic acid which comprises it. Primers are chosen to selectively hybridize, under stringent hybridization conditions, to a polynucleotide of the present invention. Generally, the primers are complementary to a subsequence of the target nucleic acid which they amplify. As those skilled in the art will appreciate, the sites to which the primer pairs will selectively hybridize are chosen such that a single contiguous nucleic acid can be formed under the desired amplification conditions.

In optional embodiments, the primers will be constructed so that they selectively hybridize under stringent conditions to a sequence (or its complement) within the target nucleic acid which comprises the codon encoding the carboxy or amino terminal amino acid residue (i.e., the 3' terminal coding region and 5' terminal coding region, respectively) of the polynucleotides of the present invention. Optionally within these embodiments, the primers will be constructed to selectively hybridize entirely within the coding region of the target polynucleotide of the present invention such that the product of amplification of a cDNA target will consist of the coding region of that cDNA. The primer length in nucleotides is selected from the group of integers consisting of from at least 15 to 50. Thus, the primers can be at least 15, 18, 20, 25, 30, 40, or 50 nucleotides in length. Those of skill will recognize that a lengthened primer sequence can be employed to increase specificity of binding (i.e., annealing) to a target sequence. A non-annealing sequence at the 5'end of a primer (a "tail") can be added, for example, to introduce a cloning site at the terminal ends of the amplicon.

The amplification products can be translated using expression systems well known to those of skill in the art and as discussed, *infra*. The resulting translation products can be confirmed as polypeptides of the present invention by, for example, assaying for the appropriate catalytic activity (e.g., specific activity and/or substrate specificity), or verifying the presence of one or more epitopes which are specific to a polypeptide of the present invention. Methods for protein synthesis from PCR derived templates are known in the art and available commercially. See, e.g., *Amersham Life Sciences, Inc, Catalog* '97, p.354.

Methods for obtaining 5' and/or 3' ends of a vector insert are well known in the art. See, e.g., RACE (Rapid Amplification of Complementary Ends) as described in Frohman, M. A., in *PCR Protocols: A Guide to Methods and Applications*, M. A. Innis, D. H. Gelfand, J. J. Sninsky, T. J. White, Eds. (Academic Press, Inc., San Diego), pp. 28-38 (1990)); see also, U.S. Pat. No. 5,470,722, and *Current Protocols in Molecular Biology*, Unit 15.6, Ausubel, *et al.*, Eds., Greene Publishing and Wiley-Interscience, New York (1995); Frohman and Martin, *Techniques* 1:165 (1989).

C. Polynucleotides Which Selectively Hybridize to a Polynucleotide of (A) or (B)

As indicated in (c), above, the present invention provides isolated nucleic acids comprising polynucleotides of the present invention, wherein the polynucleotides selectively hybridize, under selective hybridization conditions, to a polynucleotide of sections (A) or (B) as discussed above. Thus, the polynucleotides of this embodiment can be used for isolating, detecting, and/or quantifying nucleic acids comprising the polynucleotides of (A) or (B). For example, polynucleotides of the present invention can be used to identify, isolate, or amplify partial or full-length clones in a deposited library. In some embodiments, the polynucleotides are genomic or cDNA sequences isolated or otherwise complementary to a cDNA from a dicot or monocot nucleic acid library. Exemplary species of monocots and dicots include, but are not limited to: maize, canola, soybean, cotton, wheat, sorghum, sunflower, oats, sugar cane, millet, barley, alfalfa, and rice. Optionally, the cDNA library comprises at least 80% full-length sequences, preferably at least 85% or 90% full-length sequences, and more preferably at least 95% full-length sequences. The cDNA libraries can be normalized to increase the representation of rare sequences. Low stringency hybridization conditions are typically, but not exclusively, employed with sequences having a reduced sequence identity relative

to complementary sequences. Moderate and high stringency conditions can optionally be employed for sequences of greater identity. Low stringency conditions allow selective hybridization of sequences having about 70% sequence identity and can be employed to identify orthologous or paralogous sequences.

D. Polynucleotides Having a Specific Sequence Identity with the Polynucleotides of (A), (B) or (C)

As indicated in (d), above, the present invention provides isolated nucleic acids comprising polynucleotides of the present invention, wherein the polynucleotides have a specified identity at the nucleotide level to a polynucleotide as disclosed in sections (A), (B), or (C), above. Identity can be calculated using, for example, the BLAST, CLUSTALW, or GAP algorithms under default conditions. The percentage of identity to a reference sequence is at least 50% and, rounded upwards to the nearest integer, can be expressed as an integer selected from the group of integers consisting of from 50 to 99. Thus, for example, the percentage of identity to a reference sequence can be at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, or 95%.

Optionally, the polynucleotides of this embodiment will encode a polypeptide that will share an epitope with a polypeptide encoded by the polynucleotides of sections (A), (B), or (C). Thus, these polynucleotides encode a first polypeptide which elicits production of antisera comprising antibodies which are specifically reactive to a second polypeptide encoded by a polynucleotide of (A), (B), or (C). However, the first polypeptide does not bind to antisera raised against itself when the antisera has been fully immunosorbed with the first polypeptide. Hence, the polynucleotides of this embodiment can be used to generate antibodies for use in, for example, the screening of expression libraries for nucleic acids comprising polynucleotides of (A), (B), or (C), or for purification of, or in immunoassays for, polypeptides encoded by the polynucleotides of (A), (B), or (C). The polynucleotides of this embodiment embrace nucleic acid sequences which can be employed for selective hybridization to a polynucleotide encoding a polypeptide of the present invention.

Screening polypeptides for specific binding to antisera can be conveniently achieved using peptide display libraries. This method involves the screening of large collections of peptides for individual members having the desired function or structure. Antibody screening of peptide display libraries is well known in the art. The displayed

peptide sequences can be from 3 to 5000 or more amino acids in length, frequently from 5-100 amino acids long, and often from about 8 to 15 amino acids long. In addition to direct chemical synthetic methods for generating peptide libraries, several recombinant DNA methods have been described. One type involves the display of a peptide sequence on the surface of a bacteriophage or cell. Each bacteriophage or cell contains the nucleotide sequence encoding the particular displayed peptide sequence. Such methods are described in PCT patent publication Nos. 91/17271, 91/18980, 91/19818, and 93/08278. Other systems for generating libraries of peptides have aspects of both *in vitro* chemical synthesis and recombinant methods. See, PCT Patent publication Nos. 92/05258, 92/14843, and 96/19256. See also, U.S. Patent Nos. 5,658,754; and 5,643,768. Peptide display libraries, vectors, and screening kits are commercially available from such suppliers as Invitrogen (Carlsbad, CA).

E. Polynucleotides Encoding a Protein Having a Subsequence from a Prototype Polypeptide and is Cross-Reactive to the Prototype Polypeptide

As indicated in (e), above, the present invention provides isolated nucleic acids comprising polynucleotides of the present invention, wherein the polynucleotides encode a protein having a subsequence of contiguous amino acids from a prototype polypeptide of the present invention such as are provided in (a), above. The length of contiguous amino acids from the prototype polypeptide is selected from the group of integers consisting of from at least 10 to the number of amino acids within the prototype sequence. Thus, for example, the polynucleotide can encode a polypeptide having a subsequence having at least 10, 15, 20, 25, 30, 35, 40, 45, or 50, contiguous amino acids from the prototype polypeptide. Further, the number of such subsequences encoded by a polynucleotide of the instant embodiment can be any integer selected from the group consisting of from 1 to 20, such as 2, 3, 4, or 5. The subsequences can be separated by any integer of nucleotides from 1 to the number of nucleotides in the sequence such as at least 5, 10, 15, 25, 50, 100, or 200 nucleotides.

The proteins encoded by polynucleotides of this embodiment, when presented as an immunogen, elicit the production of polyclonal antibodies which specifically bind to a prototype polypeptide such as but not limited to, a polypeptide encoded by the polynucleotide of (a) or (b), above. Generally, however, a protein encoded by a polynucleotide of this embodiment does not bind to antisera raised against the prototype

polypeptide when the antisera has been fully immunosorbed with the prototype polypeptide. Methods of making and assaying for antibody binding specificity/affinity are well known in the art. Exemplary immunoassay formats include ELISA, competitive immunoassays, radioimmunoassays, Western blots, indirect immunofluorescent assays and the like.

In a preferred assay method, fully immunosorbed and pooled antisera which is elicited to the prototype polypeptide can be used in a competitive binding assay to test the protein. The concentration of the prototype polypeptide required to inhibit 50% of the binding of the antisera to the prototype polypeptide is determined. If the amount of the protein required to inhibit binding is less than twice the amount of the prototype protein, then the protein is said to specifically bind to the antisera elicited to the immunogen. Accordingly, the proteins of the present invention embrace allelic variants, conservatively modified variants, and minor recombinant modifications to a prototype polypeptide.

A polynucleotide of the present invention optionally encodes a protein having a molecular weight as the non-glycosylated protein within 20% of the molecular weight of the full-length non-glycosylated polypeptides of the present invention. Molecular weight can be readily determined by SDS-PAGE under reducing conditions. Optionally, the molecular weight is within 15% of a full length polypeptide of the present invention, more preferably within 10% or 5%, and most preferably within 3%, 2%, or 1% of a full length polypeptide of the present invention.

Optionally, the polynucleotides of this embodiment will encode a protein having a specific enzymatic activity at least 50%, 60%, 80%, or 90% of a cellular extract comprising the native, endogenous full-length polypeptide of the present invention. Further, the proteins encoded by polynucleotides of this embodiment will optionally have a substantially similar affinity constant (K_m) and/or catalytic activity (i.e., the microscopic rate constant, k_{cat}) as the native endogenous, full-length protein. Those of skill in the art will recognize that k_{cat}/K_m value determines the specificity for competing substrates and is often referred to as the specificity constant. Proteins of this embodiment can have a k_{cat}/K_m value at least 10% of a full-length polypeptide of the present invention as determined using the endogenous substrate of that polypeptide. Optionally, the k_{cat}/K_m value will be at least 20%, 30%, 40%, 50%, and most preferably at least 60%, 70%, 80%, 90%, or 95% the k_{cat}/K_m value of the full-length polypeptide of the present invention. Determination of k_{cat}, K_m , and k_{cat}/K_m can be determined by any number of means well

known to those of skill in the art. For example, the initial rates (i.e., the first 5% or less of the reaction) can be determined using rapid mixing and sampling techniques (e.g., continuous-flow, stopped-flow, or rapid quenching techniques), flash photolysis, or relaxation methods (e.g., temperature jumps) in conjunction with such exemplary methods of measuring as spectrophotometry, spectrofluorimetry, nuclear magnetic resonance, or radioactive procedures. Kinetic values are conveniently obtained using a Lineweaver-Burk or Eadie-Hofstee plot.

F. Polynucleotides Complementary to the Polynucleotides of (A)-(E)

As indicated in (f), above, the present invention provides isolated nucleic acids comprising polynucleotides complementary to the polynucleotides of paragraphs A-E, above. As those of skill in the art will recognize, complementary sequences base-pair throughout the entirety of their length with the polynucleotides of sections (A)-(E) (i.e., have 100% sequence identity over their entire length). Complementary bases associate through hydrogen bonding in double stranded nucleic acids. For example, the following base pairs are complementary: guanine and cytosine; adenine and thymine; and adenine and uracil.

G. Polynucleotides Which are Subsequences of the Polynucleotides of (A)-(F)

As indicated in (g), above, the present invention provides isolated nucleic acids comprising polynucleotides which comprise at least 15 contiguous bases from the polynucleotides of sections (A) through (F) as discussed above. The length of the polynucleotide is given as an integer selected from the group consisting of from at least 15 to the length of the nucleic acid sequence from which the polynucleotide is a subsequence of. Thus, for example, polynucleotides of the present invention are inclusive of polynucleotides comprising at least 15, 20, 25, 30, 40, 50, 60, 75, or 100 contiguous nucleotides in length from the polynucleotides of (A)-(F). Optionally, the number of such subsequences encoded by a polynucleotide of the instant embodiment can be any integer selected from the group consisting of from 1 to 20, such as 2, 3, 4, or 5. The subsequences can be separated by any integer of nucleotides from 1 to the number of nucleotides in the sequence such as at least 5, 10, 15, 25, 50, 100, or 200 nucleotides.

The subsequences of the present invention can comprise structural characteristics of the sequence from which it is derived. Alternatively, the subsequences can lack certain

structural characteristics of the larger sequence from which it is derived such as a poly (A) tail. Optionally, a subsequence from a polynucleotide encoding a polypeptide having at least one epitope in common with a prototype polypeptide sequence as provided in (a), above, may encode an epitope in common with the prototype sequence. Alternatively, the subsequence may not encode an epitope in common with the prototype sequence but can be used to isolate the larger sequence by, for example, nucleic acid hybridization with the sequence from which it's derived. Subsequences can be used to modulate or detect gene expression by introducing into the subsequences compounds which bind, intercalate, cleave and/or crosslink to nucleic acids. Exemplary compounds include acridine, psoralen, phenanthroline, naphthoquinone, daunomycin or chloroethylaminoaryl conjugates.

Construction of Nucleic Acids

The isolated nucleic acids of the present invention can be made using (a) standard recombinant methods, (b) synthetic techniques, or combinations thereof. In some embodiments, the polynucleotides of the present invention will be cloned, amplified, or otherwise constructed from a monocot. In preferred embodiments the monocot is *Zea mays*.

The nucleic acids may conveniently comprise sequences in addition to a polynucleotide of the present invention. For example, a multi-cloning site comprising one or more endonuclease restriction sites may be inserted into the nucleic acid to aid in isolation of the polynucleotide. Also, translatable sequences may be inserted to aid in the isolation of the translated polynucleotide of the present invention. For example, a hexa-histidine marker sequence provides a convenient means to purify the proteins of the present invention. A polynucleotide of the present invention can be attached to a vector, adapter, or linker for cloning and/or expression of a polynucleotide of the present invention. Additional sequences may be added to such cloning and/or expression sequences to optimize their function in cloning and/or expression, to aid in isolation of the polynucleotide, or to improve the introduction of the polynucleotide into a cell. Typically, the length of a nucleic acid of the present invention less the length of its polynucleotide of the present invention is less than 20 kilobase pairs, often less than 15 kb, and frequently less than 10 kb. Use of cloning vectors, expression vectors, adapters, and linkers is well known and extensively described in the art. For a description of various nucleic acids see,

for example, *Stratagene Cloning Systems, Catalogs* 1995, 1996, 1997 (La Jolla, CA); and, *Amersham Life Sciences, Inc, Catalog* '97 (Arlington Heights, IL).

A. Recombinant Methods for Constructing Nucleic Acids

The isolated nucleic acid compositions of this invention, such as RNA, cDNA, genomic DNA, or a hybrid thereof, can be obtained from plant biological sources using any number of cloning methodologies known to those of skill in the art. In some embodiments, oligonucleotide probes which selectively hybridize, under stringent conditions, to the polynucleotides of the present invention are used to identify the desired sequence in a cDNA or genomic DNA library. While isolation of RNA, and construction of cDNA and genomic libraries is well known to those of ordinary skill in the art, the following highlights some of the methods employed.

A1. mRNA Isolation and Purification

Total RNA from plant cells comprises such nucleic acids as mitochondrial RNA, chloroplastic RNA, rRNA, tRNA, hnRNA and mRNA. Total RNA preparation typically involves lysis of cells and removal of organelles and proteins, followed by precipitation of nucleic acids. Extraction of total RNA from plant cells can be accomplished by a variety of means. Frequently, extraction buffers include a strong detergent such as SDS and an organic denaturant such as guanidinium isothiocyanate, guanidine hydrochloride or phenol. Following total RNA isolation, poly(A)$^+$ mRNA is typically purified from the remainder RNA using oligo(dT) cellulose. Exemplary total RNA and mRNA isolation protocols are described in *Plant Molecular Biology: A Laboratory Manual*, Clark, Ed., Springer-Verlag, Berlin (1997); and, *Current Protocols in Molecular Biology*, Ausubel, *et al.*, Eds., Greene Publishing and Wiley-Interscience, New York (1995). Total RNA and mRNA isolation kits are commercially available from vendors such as Stratagene (La Jolla, CA), Clonetech (Palo Alto, CA), Pharmacia (Piscataway, NJ), and 5'-3' (Paoli Inc., PA). See also, U.S. Patent Nos. 5,614,391; and, 5,459,253. The mRNA can be fractionated into populations with size ranges of about 0.5, 1.0, 1.5, 2.0, 2.5 or 3.0 kb. The cDNA synthesized for each of these fractions can be size selected to the same size range as its mRNA prior to vector insertion. This method helps eliminate truncated cDNA formed by incompletely reverse transcribed mRNA.

A2. Construction of a cDNA Library

Construction of a cDNA library generally entails five steps. First, first strand cDNA synthesis is initiated from a poly(A)$^+$ mRNA template using a poly(dT) primer or random hexanucleotides. Second, the resultant RNA-DNA hybrid is converted into double stranded cDNA, typically by reaction with a combination of RNAse H and DNA polymerase I (or Klenow fragment). Third, the termini of the double stranded cDNA are ligated to adaptors. Ligation of the adaptors can produce cohesive ends for cloning. Fourth, size selection of the double stranded cDNA eliminates excess adaptors and primer fragments, and eliminates partial cDNA molecules due to degradation of mRNAs or the failure of reverse transcriptase to synthesize complete first strands. Fifth, the cDNAs are ligated into cloning vectors and packaged. cDNA synthesis protocols are well known to the skilled artisan and are described in such standard references as: *Plant Molecular Biology: A Laboratory Manual*, Clark, Ed., Springer-Verlag, Berlin (1997); and, *Current Protocols in Molecular Biology*, Ausubel, *et al.*, Eds., Greene Publishing and Wiley-Interscience, New York (1995). cDNA synthesis kits are available from a variety of commercial vendors such as Stratagene or Pharmacia.

A number of cDNA synthesis protocols have been described which provide substantially pure full-length cDNA libraries. Substantially pure full-length cDNA libraries are constructed to comprise at least 90%, and more preferably at least 93% or 95% full-length inserts amongst clones containing inserts. The length of insert in such libraries can be from 0 to 8, 9, 10, 11, 12, 13, or more kilobase pairs. Vectors to accommodate inserts of these sizes are known in the art and available commercially. See, e.g., Stratagene's lambda ZAP Express (cDNA cloning vector with 0 to 12 kb cloning capacity).

An exemplary method of constructing a greater than 95% pure full-length cDNA library is described by Carninci *et al.*, *Genomics*, 37:327-336 (1996). In that protocol, the cap-structure of eukaryotic mRNA is chemically labeled with biotin. By using streptavidin-coated magnetic beads, only the full-length first-strand cDNA/mRNA hybrids are selectively recovered after RNase I treatment. The method provides a high yield library with an unbiased representation of the starting mRNA population. Other methods for producing full-length libraries are known in the art. See, e.g., Edery *et al.*, *Mol. Cell Biol.*,15(6):3363-3371 (1995); and, PCT Application WO 96/34981.

A3. Normalized or Subtracted cDNA Libraries

A non-normalized cDNA library represents the mRNA population of the tissue it was made from. Since unique clones are out-numbered by clones derived from highly expressed genes their isolation can be laborious. Normalization of a cDNA library is the process of creating a library in which each clone is more equally represented.

A number of approaches to normalize cDNA libraries are known in the art. One approach is based on hybridization to genomic DNA. The frequency of each hybridized cDNA in the resulting normalized library would be proportional to that of each corresponding gene in the genomic DNA. Another approach is based on kinetics. If cDNA reannealing follows second-order kinetics, rarer species anneal less rapidly and the remaining single-stranded fraction of cDNA becomes progressively more normalized during the course of the hybridization. Specific loss of any species of cDNA, regardless of its abundance, does not occur at any Cot value. Construction of normalized libraries is described in Ko, *Nucl. Acids. Res.*, 18(19):5705-5711 (1990); Patanjali *et al.*, *Proc. Nat'l. Acad. U.S.A.*, 88:1943-1947 (1991); U.S. Patents 5,482,685, and 5,637,685. In an exemplary method described by Soares *et al.*, normalization resulted in reduction of the abundance of clones from a range of four orders of magnitude to a narrow range of only 1 order of magnitude. *Proc. Nat'l. Acad. Sci. USA*, 91:9228-9232 (1994).

Subtracted cDNA libraries are another means to increase the proportion of less abundant cDNA species. In this procedure, cDNA prepared from one pool of mRNA is depleted of sequences present in a second pool of mRNA by hybridization. The cDNA:mRNA hybrids are removed and the remaining un-hybridized cDNA pool is enriched for sequences unique to that pool. See, Foote *et al.* in, *Plant Molecular Biology: A Laboratory Manual*, Clark, Ed., Springer-Verlag, Berlin (1997); Kho and Zarbl, *Technique*, 3(2):58-63 (1991); Sive and St. John, *Nucl. Acids Res.*, 16(22):10937 (1988); *Current Protocols in Molecular Biology*, Ausubel, *et al.*, Eds., Greene Publishing and Wiley-Interscience, New York (1995); and, Swaroop *et al.*, *Nucl. Acids Res.*, 19(8):1954 (1991). cDNA subtraction kits are commercially available. See, e.g., PCR-Select (Clontech, Palo Alto, CA).

A4. Construction of a Genomic Library

To construct genomic libraries, large segments of genomic DNA are generated by fragmentation, e.g. using restriction endonucleases, and are ligated with vector DNA to

form concatemers that can be packaged into the appropriate vector. Methodologies to accomplish these ends, and sequencing methods to verify the sequence of nucleic acids are well known in the art. Examples of appropriate molecular biological techniques and instructions sufficient to direct persons of skill through many construction, cloning, and screening methodologies are found in Sambrook, *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory Vols. 1-3 (1989), Methods in Enzymology, Vol. 152: *Guide to Molecular Cloning Techniques*, Berger and Kimmel, Eds., San Diego: Academic Press, Inc. (1987), *Current Protocols in Molecular Biology*, Ausubel, *et al.*, Eds., Greene Publishing and Wiley-Interscience, New York (1995); *Plant Molecular Biology: A Laboratory Manual*, Clark, Ed., Springer-Verlag, Berlin (1997). Kits for construction of genomic libraries are also commercially available.

A5. Nucleic Acid Screening and Isolation Methods

The cDNA or genomic library can be screened using a probe based upon the sequence of a polynucleotide of the present invention such as those disclosed herein. Probes may be used to hybridize with genomic DNA or cDNA sequences to isolate homologous genes in the same or different plant species. Those of skill in the art will appreciate that various degrees of stringency of hybridization can be employed in the assay; and either the hybridization or the wash medium can be stringent. As the conditions for hybridization become more stringent, there must be a greater degree of complementarity between the probe and the target for duplex formation to occur. The degree of stringency can be controlled by temperature, ionic strength, pH and the presence of a partially denaturing solvent such as formamide. For example, the stringency of hybridization is conveniently varied by changing the polarity of the reactant solution through manipulation of the concentration of formamide within the range of 0% to 50%. The degree of complementarity (sequence identity) required for detectable binding will vary in accordance with the stringency of the hybridization medium and/or wash medium. The degree of complementarity will optimally be 100 percent; however, it should be understood that minor sequence variations in the probes and primers may be compensated for by reducing the stringency of the hybridization and/or wash medium.

The nucleic acids of interest can also be amplified from nucleic acid samples using amplification techniques. For instance, polymerase chain reaction (PCR) technology can be used to amplify the sequences of polynucleotides of the present invention and related

genes directly from genomic DNA or cDNA libraries. PCR and other *in vitro* amplification methods may also be useful, for example, to clone nucleic acid sequences that code for proteins to be expressed, to make nucleic acids to use as probes for detecting the presence of the desired mRNA in samples, for nucleic acid sequencing, or for other purposes. Examples of techniques sufficient to direct persons of skill through *in vitro* amplification methods are found in Berger, Sambrook, and Ausubel, as well as Mullis *et al.*, U.S. Patent No. 4,683,202 (1987); and, *PCR Protocols A Guide to Methods and Applications*, Innis *et al.*, Eds., Academic Press Inc., San Diego, CA (1990). Commercially available kits for genomic PCR amplification are known in the art. See, e.g., Advantage-GC Genomic PCR Kit (Clontech). The T4 gene 32 protein (Boehringer Mannheim) can be used to improve yield of long PCR products.

PCR-based screening methods have also been described. Wilfinger *et al.* describe a PCR-based method in which the longest cDNA is identified in the first step so that incomplete clones can be eliminated from study. *BioTechniques*, 22(3): 481-486 (1997). In that method, a primer pair is synthesized with one primer annealing to the 5' end of the sense strand of the desired cDNA and the other primer to the vector. Clones are pooled to allow large-scale screening. By this procedure, the longest possible clone is identified amongst candidate clones. Further, the PCR product is used solely as a diagnostic for the presence of the desired cDNA and does not utilize the PCR product itself. Such methods are particularly effective in combination with a full-length cDNA construction methodology, above.

Functional fragments of cell cycle genes can be identified using a variety of techniques such as restriction analysis, Southern analysis, primer extension analysis, and DNA sequence analysis. Function can also be determined by complementing yeast strains known to be mutant for cell cycle genes with maize homologs. Primer extension analysis or S1 nuclease protection analysis, for example, can be used to localize the putative start site of transcription of the cloned gene. Ausubel at pages 4.8.1 to 4.8.5; Walmsley et al., "Quantitative and Qualitative Analysis of Exogenous Gene Expression by the S1 Nuclease Protection Assay," in *Methods in Molecular Biology*, Vol. 7: *Gene Transfer and Expression*.

The general approach of such functional analysis involves subcloning DNA fragments of a genomic clone, cDNA clone, or synthesized gene sequence into an expression vector, introducing the expression vector into a heterologous host, and relying

on an assay system such as BrdU incorporation to monitor DNA synthesis and various well-established visual methods to follow cell division. Functional fragments of cell cycle proteins are identified by their ability, upon introduction to cells, to stimulate the transition from G1 to S-phase and/or from G2 to M-phase, manifested by increased DNA replication in a population of cells and by increased cell division rates, sometimes evidenced by differential cell density and/or cell shape.

Methods for generating fragments of a cDNA or genomic clone are well known. In addition, variants can be obtained, for example, by oligonucleotide-directed mutagenesis, linker-scanning mutagenesis, mutagenesis using the polymerase chain reaction, and the like. See, for example, Ausubel, pages 8.0.3 - 8.5.9. Also, see generally, McPherson (ed.), *Directed Mutagenesis: A Practical Approach*, (IRL Press, 1991). Thus, the present invention also encompasses DNA molecules comprising nucleotide sequences that have substantial sequence similarity with sequences chosen from the group consisting of SEQ ID NOS: 1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, and 49, and encode cell cycle genes.

B. Synthetic Methods for Constructing Nucleic Acids

The isolated nucleic acids of the present invention can also be prepared by direct chemical synthesis by methods such as the phosphotriester method of Narang *et al., Meth. Enzymol.* 68: 90-99 (1979); the phosphodiester method of Brown *et al., Meth. Enzymol.* 68: 109-151 (1979); the diethylphosphoramidite method of Beaucage *et al., Tetra. Lett.* 22: 1859-1862 (1981); the solid phase phosphoramidite triester method described by Beaucage and Caruthers, *Tetra. Letts.* 22(20): 1859-1862 (1981), *e.g.*, using an automated synthesizer, *e.g.*, as described in Needham-VanDevanter *et al., Nucleic Acids Res.*, 12: 6159-6168 (1984); and, the solid support method of U.S. Patent No. 4,458,066. Chemical synthesis generally produces a single stranded oligonucleotide. This may be converted into double stranded DNA by hybridization with a complementary sequence, or by polymerization with a DNA polymerase using the single strand as a template. One of skill will recognize that while chemical synthesis of DNA is best employed for sequences of about 100 bases or less, longer sequences may be obtained by the ligation of shorter sequences.

Recombinant Expression Cassettes

The present invention further provides recombinant expression cassettes comprising a nucleic acid of the present invention. A nucleic acid sequence coding for the desired polypeptide of the present invention, for example a cDNA or a genomic sequence encoding a full length polypeptide of the present invention, can be used to construct a recombinant expression cassette which can be introduced into the desired host cell. A recombinant expression cassette will typically comprise a polynucleotide of the present invention operably linked to transcriptional initiation regulatory sequences which will direct the transcription of the polynucleotide in the intended host cell, such as tissues of a transformed plant.

For example, plant expression vectors may include (1) a cloned plant gene under the transcriptional control of 5' and 3' regulatory sequences and (2) a dominant selectable marker. Such plant expression vectors may also contain, if desired, a promoter regulatory region (e.g., one conferring inducible or constitutive, environmentally- or developmentally-regulated, or cell- or tissue-specific/selective expression), a transcription initiation start site, a ribosome binding site, an RNA processing signal, a transcription termination site, and/or a polyadenylation signal.

A plant promoter fragment can be employed which will direct expression of a polynucleotide of the present invention in all tissues of a regenerated plant. Such promoters are referred to herein as "constitutive" promoters and are active under most environmental conditions and states of development or cell differentiation. Examples of constitutive promoters include the cauliflower mosaic virus (CaMV) 35S transcription initiation region, the 1'- or 2'- promoter derived from T-DNA of *Agrobacterium tumefaciens*, the ubiquitin 1 promoter, the Smas promoter, the cinnamyl alcohol dehydrogenase promoter (U.S. Patent No. 5,683,439), the *Nos* promoter, the pEmu promoter, the rubisco promoter, the GRP1-8 promoter, and other transcription initiation regions from various plant genes known to those of skill.

Many different constitutive promoters can be utilized in the instant invention. Exemplary constitutive promoters include the promoters from plant viruses such as the 35S promoter from CaMV (Odell *et al.*, *Nature* 313: 810-812 (1985) and the promoters from such genes as rice actin (McElroy *et al.*, *Plant Cell* 2: 163-171 (1990)); ubiquitin (Christensen *et al.*, *Plant Mol. Biol.* 12: 619-632 (1989) and Christensen *et al.*, *Plant Mol. Biol.* 18: 675-689 (1992)); pEMU (Last *et al.*, *Theor. Appl. Genet.* 81: 581-588 (1991));

MAS (Velten *et al.*, *EMBO J.* 3: 2723-2730 (1984)) and maize H3 histone (Lepetit *et al.*, *Mol. Gen. Genet.* 231: 276-285 (1992) and Atanassova *et al.*, *Plant Journal* 2(3): 291-300 (1992)).

The ALS promoter, a XbaI/NcoI fragment 5-prime to the *Brassica napus* ALS3 structural gene (or a nucleotide sequence that has substantial sequence similarity to said XbaI/NcoI fragment), represents a particularly useful constitutive promoter. Co-pending Pioneer Hi-Bred International U.S. Patent Application 08/409,297.

The expression vector comprises a constitutive promoter operably linked to a nucleotide sequence comprising cell cycle genes, or the constitutive promoter is operably linked to a nucleotide sequence encoding a signal sequence which is operably linked to a nucleotide sequence comprising cell cycle genes. The expression vector is introduced into plant cells, and presumptively transformed cells are screened for the presence of cell cycle gene products by either BrdU or cell division assays, as described herein.

Alternatively, the plant promoter can direct expression of a polynucleotide of the present invention in a specific tissue or may be otherwise under more precise environmental or developmental control. Such promoters are referred to here as "inducible" promoters. With an inducible promoter, the rate of transcription increases in response to an inducing agent. Environmental conditions that may effect transcription by inducible promoters include pathogen attack, anaerobic conditions, or the presence of light. Examples of inducible promoters are the Adh1 promoter which is inducible by hypoxia or cold stress, the Hsp70 promoter which is inducible by heat stress, and the PPDK promoter which is inducible by light.

A variety of inducible promoters can be used in the instant invention. *See* Ward *et al. Plant Mol. Biol.* 22: 361-366 (1993). Exemplary inducible promoters include that from the ACE1 system which responds to copper (Mett *et al. PNAS* 90: 4567-4571 (1993)); In2 gene from maize which responds to benzenesulfonamide herbicide safeners (Hershey *et al.*, *Mol. Gen. Genetics* 227: 229-237 (1991) and Gatz *et al.*, *Mol. Gen. Genetics* 243: 32-38 (1994)); or Tet repressor from Tn10 (Gatz *et al.*, *Mol. Gen. Genet.* 227: 229-237 (1991). A particularly preferred inducible promoter is a promoter that responds to an inducing agent to which plants do not normally respond. An exemplary inducible promoter is the inducible promoter from a steroid hormone gene the transcriptional activity of which is induced by a glucocorticosteroid hormone. Schena *et al.*, *Proc. Nat'l. Acad. Sci. U.S.A.* 88: 10421 (1991).

An inducible promoter is operably linked to a nucleotide sequence comprising cell cycle genes. Optionally, the inducible promoter is operably linked to a nucleotide sequence encoding a signal sequence which is operably linked to a nucleotide sequence comprising cell cycle genes. The expression vector is introduced into plant cells and presumptively transformed cells are exposed to an inducer of the inducible promoter. The cells are screened for the presence of cell cycle proteins by either BrdU or cell division assays, as described herein.

Examples of promoters under developmental control include promoters that initiate transcription only, or preferentially, in certain tissues, such as leaves, roots, fruit, seeds, or flowers. Exemplary promoters include the anther specific promoter 5126 (U.S. Patent Nos. 5,689,049 and 5,689,051), glob-1 promoter, and gamma-zein promoter. The operation of a promoter may also vary depending on its location in the genome. Thus, an inducible promoter may become fully or partially constitutive in certain locations.

A variety of tissue-specific or tissue-preferred promoters can be utilized in the instant invention. Exemplary tissue-specific or tissue-preferred promoters include a seed-preferred promoter such as that from the phaseolin gene (Murai *et al.*, *Science* 23: 476-482 (1983) and Sengupta-Gopalan *et al.*, *Proc. Nat'l. Acad. Sci. USA* 82: 3320-3324 (1985)); a leaf-specific and light-induced promoter such as that from *cab* or *rubisco* (Simpson *et al.*, *EMBO J.* 4(11): 2723-2729 (1985) and Timko *et al.*, *Nature* 318: 579-582 (1985)); an anther-specific promoter such as that from *LAT52* (Twell *et al.*, *Mol. Gen. Genet.* 217: 240-245 (1989)); a pollen-specific promoter such as that from *Zm13* (Guerrero *et al.*, *Mol. Gen. Genet.* 224: 161-168 (1993)) or a microspore-preferred promoter such as that from *apg* (Twell *et al.*, *Sex. Plant Reprod.* 6: 217-224 (1993).

The expression vector comprises a tissue-specific or tissue-preferred promoter operably linked to a nucleotide sequence comprising cell cycle genes. Optionally, the tissue-specific promoter is operably linked to a nucleotide sequence encoding a signal sequence which is operably linked to a nucleotide sequence comprising cell cycle genes. The expression vector is introduced into plant cells. The cells are screened for the presence of cell cycle proteins by either BrdU or cell division assays, as described herein. Plants transformed with cell cycle genes operably linked to a tissue-specific promoter produce the cell cycle proteins exclusively, or preferentially, in a specific tissue.

Both heterologous and non-heterologous (i.e., endogenous) promoters can be employed to direct expression of the nucleic acids of the present invention. These

promoters can also be used, for example, in recombinant expression cassettes to drive expression of antisense nucleic acids to reduce, increase, or alter concentration and/or composition of the proteins of the present invention in a desired tissue. Thus, in some embodiments, the nucleic acid construct will comprise a promoter functional in a plant cell, such as in *Zea mays*, operably linked to a polynucleotide of the present invention. Promoters useful in these embodiments include the endogenous promoters driving expression of a polypeptide of the present invention.

In some embodiments, isolated nucleic acids which serve as promoter or enhancer elements can be introduced in the appropriate position (generally upstream) of a non-heterologous form of a polynucleotide of the present invention so as to up or down regulate expression of a polynucleotide of the present invention. For example, endogenous promoters can be altered *in vivo* by mutation, deletion, and/or substitution (see, Kmiec, U.S. Patent 5,565,350; Zarling *et al.*, WO 93/22443), or isolated promoters can be introduced into a plant cell in the proper orientation and distance from a gene of the present invention so as to control the expression of the gene. Gene expression can be modulated under conditions suitable for plant growth so as to alter the total concentration and/or alter the composition of the polypeptides of the present invention in plant cell. Thus, the present invention provides compositions, and methods for making, heterologous promoters and/or enhancers operably linked to a native, endogenous (i.e., non-heterologous) form of a polynucleotide of the present invention.

Methods for identifying promoters with a particular expression pattern, in terms of, e.g., tissue type, cell type, stage of development, and/or environmental conditions, are well known in the art. See, e.g., *The Maize Handbook*, Chapters 114-115, Freeling and Walbot, Eds., Springer, New York (1994); *Corn and Corn Improvement*, 3rd edition, Chapter 6, Sprague and Dudley, Eds., American Society of Agronomy, Madison, Wisconsin (1988). A typical step in promoter isolation methods is identification of gene products that are expressed with some degree of specificity in the target tissue. Amongst the range of methodologies are: differential hybridization to cDNA libraries; subtractive hybridization; differential display; differential 2-D protein gel electrophoresis; DNA probe arrays; and isolation of proteins known to be expressed with some specificity in the target tissue. Such methods are well known to those of skill in the art. Commercially available products for identifying promoters are known in the art such as Clontech's (Palo Alto, CA) Universal GenomeWalker Kit.

For the protein-based methods, it is helpful to obtain the amino acid sequence for at least a portion of the identified protein, and then to use the protein sequence as the basis for preparing a nucleic acid that can be used as a probe to identify either genomic DNA directly, or preferably, to identify a cDNA clone from a library prepared from the target tissue. Once such a cDNA clone has been identified, that sequence can be used to identify the sequence at the 5' end of the transcript of the indicated gene. For differential hybridization, subtractive hybridization and differential display, the nucleic acid sequence identified as enriched in the target tissue is used to identify the sequence at the 5' end of the transcript of the indicated gene. Once such sequences are identified, starting either from protein sequences or nucleic acid sequences, any of these sequences identified as being from the gene transcript can be used to screen a genomic library prepared from the target organism. Methods for identifying and confirming the transcriptional start site are well known in the art.

In the process of isolating promoters expressed under particular environmental conditions or stresses, or in specific tissues, or at particular developmental stages, a number of genes are identified that are expressed under the desired circumstances, in the desired tissue, or at the desired stage. Further analysis will reveal expression of each particular gene in one or more other tissues of the plant. One can identify a promoter with activity in the desired tissue or condition but that does not have activity in any other common tissue.

To identify the promoter sequence, the 5' portions of the clones described here are analyzed for sequences characteristic of promoter sequences. For instance, promoter sequence elements include the TATA box consensus sequence (TATAAT), which is usually an AT-rich stretch of 5-10 bp located approximately 20 to 40 base pairs upstream of the transcription start site. Identification of the TATA box is well known in the art. For example, one way to predict the location of this element is to identify the transcription start site using standard RNA-mapping techniques such as primer extension, S1 analysis, and/or RNase protection. To confirm the presence of the AT-rich sequence, a structure-function analysis can be performed involving mutagenesis of the putative region and quantification of the mutation's effect on expression of a linked downstream reporter gene. See, e.g., *The Maize Handbook*, Chapter 114, Freeling and Walbot, Eds., Springer, New York, (1994).

In plants, further upstream from the TATA box, at positions -80 to -100, there is typically a promoter element (i.e., the CAAT box) with a series of adenines surrounding

the trinucleotide G (or T) N G. J. Messing *et al.*, in *Genetic Engineering in Plants*, Kosage, Meredith and Hollaender, Eds., pp. 221-227, 1983. In maize, there is no well-conserved CAAT box but there are several short, conserved protein-binding motifs upstream of the TATA box. These include motifs for the trans-acting transcription factors involved in light regulation, anaerobic induction, hormonal regulation, or anthocyanin biosynthesis, as appropriate for each gene.

Once promoter and/or gene sequences are known, a region of suitable size is selected from the genomic DNA that is 5' to the transcriptional start, or the translational start site, and such sequences are then linked to a coding sequence. If the transcriptional start site is used as the point of fusion, any of a number of possible 5' untranslated regions can be used in between the transcriptional start site and the partial coding sequence. If the translational start site at the 3' end of the specific promoter is used, then it is linked directly to the methionine start codon of a coding sequence.

If polypeptide expression is desired, it is generally desirable to include a polyadenylation region at the 3'-end of a polynucleotide coding region. The polyadenylation region can be derived from the natural gene, from a variety of other plant genes, or from T-DNA. The 3' end sequence to be added can be derived from, for example, the nopaline synthase or octopine synthase genes, or alternatively from another plant gene, or less preferably from any other eukaryotic gene.

An intron sequence can be added to the 5' untranslated region or the coding sequence of the partial coding sequence to increase the amount of the mature message that accumulates in the cytosol. Inclusion of a spliceable intron in the transcription unit in both plant and animal expression constructs has been shown to increase gene expression at both the mRNA and protein levels up to 1000-fold. Buchman and Berg, *Mol. Cell Biol.* 8: 4395-4405 (1988); Callis *et al.*, *Genes Dev.* 1: 1183-1200 (1987). Such intron enhancement of gene expression is typically greatest when placed near the 5' end of the transcription unit. Use of maize introns Adh1-S intron 1, 2, and 6, the Bronze-1 intron are known in the art. See generally, *The Maize Handbook*, Chapter 116, Freeling and Walbot, Eds., Springer, New York (1994).

The vector comprising the sequences from a polynucleotide of the present invention will typically comprise a marker gene which confers a selectable phenotype on plant cells. Usually, the selectable marker gene will encode antibiotic resistance, with suitable genes including genes coding for resistance to the antibiotic spectinomycin (e.g.,

the aada gene), the streptomycin phosphotransferase (SPT) gene coding for streptomycin resistance, the neomycin phosphotransferase (NPTII) gene encoding kanamycin or geneticin resistance, the hygromycin phosphotransferase (HPT) gene coding for hygromycin resistance, genes coding for resistance to herbicides which act to inhibit the action of acetolactate synthase (ALS), in particular the sulfonylurea-type herbicides (e.g., the acetolactate synthase (ALS) gene containing mutations leading to such resistance in particular the S4 and/or Hra mutations), genes coding for resistance to herbicides which act to inhibit action of glutamine synthase, such as phosphinothricin or basta (e.g., the *bar* gene), or other such genes known in the art. The *bar* gene encodes resistance to the herbicide basta, the *nptII* gene encodes resistance to the antibiotics kanamycin and geneticin, and the ALS gene encodes resistance to the herbicide chlorsulfuron.

Typical vectors useful for expression of genes in higher plants are well known in the art and include vectors derived from the tumor-inducing (Ti) plasmid of *Agrobacterium tumefaciens* described by Rogers *et al.*, Meth. In Enzymol., 153:253-277 (1987). These vectors are plant integrating vectors in that on transformation, the vectors integrate a portion of vector DNA into the genome of the host plant. Exemplary *A. tumefaciens* vectors useful herein are plasmids pKYLX6 and pKYLX7 of Schardl *et al.*, Gene, 61:1-11 (1987) and Berger *et al.*, Proc. Nat'l. Acad. Sci. U.S.A., 86:8402-8406 (1989). Another useful vector herein is plasmid pBI101.2 that is available from Clontech Laboratories, Inc. (Palo Alto, CA).

A polynucleotide of the present invention can be expressed in either sense or anti-sense orientation as desired. It will be appreciated that control of gene expression in either sense or anti-sense orientation can have a direct impact on the observable plant characteristics. Antisense technology can be conveniently used to inhibit gene expression in plants. To accomplish this, a nucleic acid segment from the desired gene is cloned and operably linked to a promoter such that the anti-sense strand of RNA will be transcribed. The construct is then transformed into plants and the antisense strand of RNA is produced. In plant cells, it has been shown that antisense RNA inhibits gene expression by preventing the accumulation of mRNA which encodes the enzyme of interest, see, e.g., Sheehy *et al.*, *Proc. Nat'l. Acad. Sci. (USA)* 85: 8805-8809 (1988); and Hiatt *et al.*, U.S. Patent No. 4,801,340.

Another method of suppression is sense suppression. Introduction of nucleic acid configured in the sense orientation has been shown to be an effective means by which to

block the transcription of target genes. For an example of the use of this method to modulate expression of endogenous genes see, Napoli *et al.*, *The Plant Cell* 2: 279-289 (1990) and U.S. Patent No. 5,034,323.

Catalytic RNA molecules or ribozymes can also be used to inhibit expression of plant genes. It is possible to design ribozymes that specifically pair with virtually any target RNA and cleave the phosphodiester backbone at a specific location, thereby functionally inactivating the target RNA. In carrying out this cleavage, the ribozyme is not itself altered, and is thus capable of recycling and cleaving other molecules, making it a true enzyme. The inclusion of ribozyme sequences within antisense RNAs confers RNA-cleaving activity upon them, thereby increasing the activity of the constructs. The design and use of target RNA-specific ribozymes is described in Haseloff *et al.*, *Nature* 334: 585-591 (1988).

A variety of cross-linking agents, alkylating agents and radical generating species as pendant groups on polynucleotides of the present invention can be used to bind, label, detect, and/or cleave nucleic acids. For example, Vlassov, V. V., *et al.*, *Nucleic Acids Research* (1986) 14:4065-4076, describe covalent bonding of a single-stranded DNA fragment with alkylating derivatives of nucleotides complementary to target sequences. A report of similar work by the same group is that by Knorre, D. G., *et al.*, *Biochimie* (1985) 67:785-789. Iverson and Dervan also showed sequence-specific cleavage of single-stranded DNA mediated by incorporation of a modified nucleotide which was capable of activating cleavage (*J Am Chem Soc* (1987) 109:1241-1243). Meyer, R. B., *et al.*, *J Am Chem Soc* (1989) 111:8517-8519, effect covalent crosslinking to a target nucleotide using an alkylating agent complementary to the single-stranded target nucleotide sequence. A photoactivated crosslinking to single-stranded oligonucleotides mediated by psoralen was disclosed by Lee, B. L., *et al.*, *Biochemistry* (1988) 27:3197-3203. Use of crosslinking in triple-helix forming probes was also disclosed by Home, *et al.*, *J Am Chem Soc* (1990) 112:2435-2437. Use of N4, N4-ethanocytosine as an alkylating agent to crosslink to single-stranded oligonucleotides has also been described by Webb and Matteucci, *J Am Chem Soc* (1986) 108:2764-2765; *Nucleic Acids Research* (1986) 14:7661-7674; Feteritz *et al.*, *J Am Chem Soc* 113:4000 (1991). Various compounds to bind, detect, label, and/or cleave nucleic acids are known in the art. See, for example, U.S. Patent Nos. 5,543,507; 5,672,593; 5,484,908; 5,256,648; and, 5,681941.

Proteins

The isolated proteins of the present invention comprise a polypeptide having at least 10 amino acids encoded by any one of the polynucleotides of the present invention as discussed more fully, above, or polypeptides which are conservatively modified variants thereof. The proteins of the present invention or variants thereof can comprise any number of contiguous amino acid residues from a polypeptide of the present invention, wherein that number is selected from the group of integers consisting of from 10 to the number of residues in a full-length polypeptide of the present invention. Optionally, this subsequence of contiguous amino acids is at least 15, 20, 25, 30, 35, or 40 amino acids in length, often at least 50, 60, 70, 80, or 90 amino acids in length. Further, the number of such subsequences can be any integer selected from the group consisting of from 1 to 20, such as 2, 3, 4, or 5.

The present invention further provides a protein comprising a polypeptide having a specified sequence identity/similarity with a polypeptide of the present invention. The percentage of sequence identity/similarity is an integer selected from the group consisting of from 50 to 99. Exemplary sequence identity/similarity values include 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, and 95%. Sequence identity can be determined using, for example, the GAP, CLUSTALW, or BLAST algorithms.

As those of skill will appreciate, the present invention includes catalytically active polypeptides of the present invention (i.e., enzymes). Catalytically active polypeptides have a specific activity of at least 20%, 30%, or 40%, and preferably at least 50%, 60%, or 70%, and most preferably at least 80%, 90%, or 95% that of the native (non-synthetic), endogenous polypeptide. Further, the substrate specificity (k_{cat}/K_m) is optionally substantially similar to the native (non-synthetic), endogenous polypeptide. Typically, the K_m will be at least 30%, 40%, or 50%, that of the native (non-synthetic), endogenous polypeptide; and more preferably at least 60%, 70%, 80%, or 90%. Methods of assaying and quantifying measures of enzymatic activity and substrate specificity (k_{cat}/K_m), are well known to those of skill in the art.

Generally, the proteins of the present invention will, when presented as an immunogen, elicit production of an antibody specifically reactive to a polypeptide of the present invention. Further, the proteins of the present invention will not bind to antisera raised against a polypeptide of the present invention which has been fully immunosorbed with the same polypeptide. Immunoassays for determining binding are well known to

those of skill in the art. A preferred immunoassay is a competitive immunoassay as discussed, *infra*. Thus, the proteins of the present invention can be employed as immunogens for constructing antibodies immunoreactive to a protein of the present invention for such exemplary utilities as immunoassays or protein purification techniques.

Expression of Proteins in Host Cells

Using the nucleic acids of the present invention, one may express a protein of the present invention in a recombinantly engineered cell such as bacteria, yeast, insect, mammalian, or preferably plant cells. The cells produce the protein in a non-natural condition (e.g., in quantity, composition, location, and/or time), because they have been genetically altered through human intervention to do so.

It is expected that those of skill in the art are knowledgeable in the numerous expression systems available for expression of a nucleic acid encoding a protein of the present invention. No attempt to describe in detail the various methods known for the expression of proteins in prokaryotes or eukaryotes will be made.

In brief summary, the expression of isolated nucleic acids encoding a protein of the present invention will typically be achieved by operably linking, for example, the DNA or cDNA to a promoter (which is either constitutive or regulatable), followed by incorporation into an expression vector. The vectors can be suitable for replication and integration in either prokaryotes or eukaryotes. Typical expression vectors contain transcription and translation terminators, initiation sequences, and promoters useful for regulation of the expression of the DNA encoding a protein of the present invention. To obtain high level expression of a cloned gene, it is desirable to construct expression vectors which contain, at the minimum, a strong promoter to direct transcription, a ribosome binding site for translational initiation, and a transcription/translation terminator. One of skill would recognize that modifications can be made to a protein of the present invention without diminishing its biological activity. Some modifications may be made to facilitate the cloning, expression, or incorporation of the targeting molecule into a fusion protein. Such modifications are well known to those of skill in the art and include, for example, a methionine added at the amino terminus to provide an initiation site, or additional amino acids (*e.g.*, poly His) placed on either terminus to create conveniently located purification sequences. Restriction sites or termination codons can also be introduced.

A. Expression in Prokaryotes

Prokaryotic cells may be used as hosts for expression. Prokaryotes most frequently are represented by various strains of *E. coli;* however, other microbial strains may also be used. Commonly used prokaryotic control sequences which are defined herein to include promoters for transcription initiation, optionally with an operator, along with ribosome binding site sequences, include such commonly used promoters as the beta lactamase (penicillinase) and lactose (lac) promoter systems (Chang et al., *Nature* 198:1056 (1977)), the tryptophan (trp) promoter system (Goeddel et al., *Nucleic Acids Res.* 8:4057 (1980)) and the lambda derived P L promoter and N-gene ribosome binding site (Shimatake *et al.*, *Nature* 292:128 (1981)). The inclusion of selection markers in DNA vectors transfected in *E. coli* is also useful. Examples of such markers include genes specifying resistance to ampicillin, tetracycline, or chloramphenicol.

The vector is selected to allow introduction into the appropriate host cell. Bacterial vectors are typically of plasmid or phage origin. Appropriate bacterial cells are infected with phage vector particles or transfected with naked phage vector DNA. If a plasmid vector is used, the bacterial cells are transfected with the plasmid vector DNA. Expression systems for expressing a protein of the present invention are available using *Bacillus sp.* and *Salmonella* (Palva, *et al.*, *Gene* 22: 229-235 (1983); Mosbach, *et al.*, *Nature* 302: 543-545 (1983)).

B. Expression in Eukaryotes

A variety of eukaryotic expression systems such as yeast, insect cell lines, plant and mammalian cells, are known to those of skill in the art. As explained briefly below, a polynucleotide of the present invention can be expressed in these eukaryotic systems. In some embodiments, transformed/transfected plant cells, as discussed *infra*, are employed as expression systems for production of the proteins of the instant invention.

Synthesis of heterologous proteins in yeast is well known. Sherman, F., *et al.*, *Methods in Yeast Genetics*, Cold Spring Harbor Laboratory (1982) is a well recognized work describing the various methods available to produce the protein in yeast. Two widely utilized yeast for production of eukaryotic proteins are *Saccharomyces cerevisiae* and *Pichia pastoris*. Vectors, strains, and protocols for expression in *Saccharomyces* and *Pichia* are known in the art and available from commercial suppliers (e.g., Invitrogen). Suitable vectors usually have expression control sequences, such as promoters, including

3-phosphoglycerate kinase or alcohol oxidase, and an origin of replication, termination sequences and the like as desired.

A protein of the present invention, once expressed, can be isolated from yeast by lysing the cells and applying standard protein isolation techniques to the lysates. The monitoring of the purification process can be accomplished by using Western blot techniques or radioimmunoassay of other standard immunoassay techniques.

The sequences encoding proteins of the present invention can also be ligated to various expression vectors for use in transfecting cell cultures of, for instance, mammalian, insect, or plant origin. Illustrative of cell cultures useful for the production of the peptides are mammalian cells. Mammalian cell systems often will be in the form of monolayers of cells although mammalian cell suspensions may also be used. A number of suitable host cell lines capable of expressing intact proteins have been developed in the art, and include the HEK293, BHK21, and CHO cell lines. Expression vectors for these cells can include expression control sequences, such as an origin of replication, a promoter (*e.g.*, the CMV promoter, a HSV *tk* promoter or *pgk* (phosphoglycerate kinase) promoter), an enhancer (Queen *et al., Immunol. Rev.* 89: 49 (1986)), and necessary processing information sites, such as ribosome binding sites, RNA splice sites, polyadenylation sites (*e.g.*, an SV40 large T Ag poly A addition site), and transcriptional terminator sequences. Other animal cells useful for production of proteins of the present invention are available, for instance, from the American Type Culture Collection.

Appropriate vectors for expressing proteins of the present invention in insect cells are usually derived from the SF9 baculovirus. Suitable insect cell lines include mosquito larvae, silkworm, armyworm, moth and *Drosophila* cell lines such as a Schneider cell line (See, Schneider, *J. Embryol. Exp. Morphol.* 27: 353-365 (1987).

As with yeast, when higher animal or plant host cells are employed, polyadenylation or transcription terminator sequences are typically incorporated into the vector. An example of a terminator sequence is the polyadenylation sequence from the bovine growth hormone gene. Sequences for accurate splicing of the transcript may also be included. An example of a splicing sequence is the VP1 intron from SV40 (Sprague, *et al., J. Virol.* 45: 773-781 (1983)). Additionally, gene sequences to control replication in the host cell may be incorporated into the vector such as those found in bovine papilloma virus type-vectors. Saveria-Campo, M., Bovine Papilloma Virus DNA a Eukaryotic

Cloning Vector in *DNA Cloning Vol. II a Practical Approach,* D.M. Glover, Ed., IRL Press, Arlington, Virginia pp. 213-238 (1985).

Transfection/Transformation of Cells

The method of transformation/transfection is not critical to the instant invention; various methods of transformation or transfection are currently available. As newer methods are available to transform crops or other host cells they may be directly applied. Accordingly, a wide variety of methods have been developed to insert a DNA sequence into the genome of a host cell to obtain the transcription and/or translation of the sequence to effect phenotypic changes in the organism. Thus, any method which provides for effective transformation/transfection may be employed.

A. Plant Transformation

A DNA sequence coding for the desired polypeptide of the present invention, for example a cDNA or a genomic sequence encoding a full length protein, will be used to construct a recombinant expression cassette which can be introduced into the desired plant.

Isolated nucleic acid acids of the present invention can be introduced into plants according to techniques known in the art. Generally, recombinant expression cassettes as described above and suitable for transformation of plant cells are prepared. Techniques for transforming a wide variety of higher plant species are well known and described in the technical, scientific, and patent literature. See, for example, Weising *et al., Ann. Rev. Genet.* 22: 421-477 (1988). For example, the DNA construct may be introduced directly into the genomic DNA of the plant cell using techniques such as electroporation, polyethylene glycol (PEG), poration, particle bombardment, silicon fiber delivery, or microinjection of plant cell protoplasts or embryogenic callus. See, e.g., Tomes, *et al.*, Direct DNA Transfer into Intact Plant Cells Via Microprojectile Bombardment. pp.197-213 in Plant Cell, Tissue and Organ Culture, Fundamental Methods. eds. O. L. Gamborg and G.C. Phillips. Springer-Verlag Berlin Heidelberg New York, 1995. Alternatively, the DNA constructs may be combined with suitable T-DNA flanking regions and introduced into a conventional *Agrobacterium tumefaciens* host vector. The virulence functions of the *Agrobacterium tumefaciens* host will direct the insertion of the construct and adjacent marker into the plant cell DNA when the cell is infected by the bacteria. See, U.S. Patent No. 5,591,616.

The introduction of DNA constructs using PEG precipitation is described in Paszkowski *et al., Embo J.* 3: 2717-2722 (1984). Electroporation techniques are described in Fromm *et al., Proc. Nat'l. Acad. Sci. (USA)* 82: 5824 (1985). Ballistic transformation techniques are described in Klein *et al., Nature* 327: 70-73 (1987), and Tomes et al., U.S. Patent No. 5,886,244. *Agrobacterium tumefaciens*-mediated transformation techniques are well described in the scientific literature. See, for example Horsch *et al., Science* 233: 496-498 (1984), and Fraley *et al., Proc. Nat'l. Acad. Sci. (USA)* 80: 4803 (1983). Although *Agrobacterium* is useful primarily in dicots, certain monocots can be transformed by *Agrobacterium*. For instance, *Agrobacterium* transformation of maize is described in U.S. Patent No. 5,550,318.

Other methods of transfection or transformation include (1) *Agrobacterium rhizogenes*-mediated transformation (see, e.g., Lichtenstein and Fuller In: Genetic Engineering, vol. 6, PWJ Rigby, Ed., London, Academic Press, 1987; and Lichtenstein, C. P., and Draper, J,. In: DNA Cloning, Vol. II, D. M. Glover, Ed.. Oxford, IRI Press, 1985), Application PCT/US87/02512 (WO 88/02405 published Apr. 7, 1988) describes the use of *A. rhizogenes* strain A4 and its Ri plasmid along with *A. tumefaciens* vectors pARC8 or pARC16 (2) liposome-mediated DNA uptake (see, e.g., Freeman *et al., Plant Cell Physiol.* 25:1353 (1984)), (3) the vortexing method (see, e.g., Kindle, *Proc. Nat'l. Acad. Sci., (USA)* 87: 1228 (1990).

DNA can also be introduced into plants by direct DNA transfer into pollen as described by Zhou *et al.*, Methods in Enzymology, 101:433 (1983); D. Hess, *Intern Rev. Cytol.*, 107:367 (1987); Luo *et al., Plant Mol. Biol. Reporter*, 6:165 (1988). Expression of polypeptide coding genes can be obtained by injection of the DNA into reproductive organs of a plant as described by Pena *et al., Nature*, 325.:274 (1987). DNA can also be injected directly into the cells of immature embryos and the rehydration of desiccated embryos as described by Neuhaus *et al., Theor. Appl. Genet.*, 75:30 (1987); and Benbrook *et al., in Proceedings Bio Expo* 1986, Butterworth, Stoneham, Mass., pp. 27-54 (1986). A variety of plant viruses that can be employed as vectors are known in the art and include cauliflower mosaic virus (CaMV), geminivirus, brome mosaic virus, and tobacco mosaic virus.

B. Transfection of Prokaryotes, Lower Eukaryotes, and Animal Cells

Animal and lower eukaryotic (e.g., yeast) host cells are competent or rendered competent for transfection by various means. There are several well-known methods of introducing DNA into animal cells. These include: calcium phosphate precipitation, fusion of the recipient cells with bacterial protoplasts containing the DNA, treatment of the recipient cells with liposomes containing the DNA, DEAE dextran, electroporation, biolistics, and micro-injection of the DNA directly into the cells. The transfected cells are cultured by means well known in the art. Kuchler, R.J., *Biochemical Methods in Cell Culture and Virology*, Dowden, Hutchinson and Ross, Inc. (1977).

Synthesis of Proteins

The proteins of the present invention can be constructed using non-cellular synthetic methods. Solid phase synthesis of proteins of less than about 50 amino acids in length may be accomplished by attaching the C-terminal amino acid of the sequence to an insoluble support followed by sequential addition of the remaining amino acids in the sequence. Techniques for solid phase synthesis are described by Barany and Merrifield, Solid-Phase Peptide Synthesis, pp. 3-284 in *The Peptides: Analysis, Synthesis, Biology. Vol. 2: Special Methods in Peptide Synthesis, Part A.*; Merrifield, *et al., J. Am. Chem. Soc.* 85: 2149-2156 (1963), and *Stewart et al., Solid Phase Peptide Synthesis, 2nd ed.*, Pierce Chem. Co., Rockford, Ill. (1984). Proteins of greater length may be synthesized by condensation of the amino and carboxy termini of shorter fragments. Methods of forming peptide bonds by activation of a carboxy terminal end (e.g., by the use of the coupling reagent N,N'-dicycylohexylcarbodiimide) are known to those of skill.

Purification of Proteins

The proteins of the present invention may be purified by standard techniques well known to those of skill in the art. Recombinantly produced proteins of the present invention can be directly expressed or expressed as a fusion protein. The recombinant protein is purified by a combination of cell lysis (e.g., sonication, French press) and affinity chromatography. For fusion products, subsequent digestion of the fusion protein with an appropriate proteolytic enzyme releases the desired recombinant protein.

The proteins of this invention, recombinant or synthetic, may be purified to substantial purity by standard techniques well known in the art, including detergent

solubilization, selective precipitation with such substances as ammonium sulfate, column chromatography, immunopurification methods, and others. *See*, for instance, R. Scopes, *Protein Purification: Principles and Practice*, Springer-Verlag: New York (1982); Deutscher, *Guide to Protein Purification*, Academic Press (1990). For example, antibodies may be raised to the proteins as described herein. Purification from *E. coli* can be achieved following procedures described in U.S. Patent No. 4,511,503. The protein may then be isolated from cells expressing the protein and further purified by standard protein chemistry techniques as described herein. Detection of the expressed protein is achieved by methods known in the art and include, for example, radioimmunoassays, Western blotting techniques or immunoprecipitation.

Transgenic Plant Regeneration

Transformed plant cells which are derived by any of the above transformation techniques can be cultured to regenerate a whole plant which possesses the transformed genotype. Such regeneration techniques often rely on manipulation of certain phytohormones in a tissue culture growth medium. For transformation and regeneration of maize see, Gordon-Kamm *et al.*, *The Plant Cell*, 2:603-618 (1990).

Plants cells transformed with a plant expression vector can be regenerated, e.g., from single cells, callus tissue or leaf discs according to standard plant tissue culture techniques. It is well known in the art that various cells, tissues, and organs from almost any plant can be successfully cultured to regenerate an entire plant. Plant regeneration from cultured protoplasts is described in Evans *et al.*, *Protoplasts Isolation and Culture*, *Handbook of Plant Cell Culture*, Macmillan Publishing Company, New York, pp. 124-176 (1983); and Binding, *Regeneration of Plants*, *Plant Protoplasts*, CRC Press, Boca Raton, pp. 21-73 (1985).

The regeneration of plants containing the foreign gene introduced by *Agrobacterium* from leaf explants can be achieved as described by Horsch *et al.*, *Science*, 227:1229-1231 (1985). In this procedure, transformants are grown in the presence of a selection agent and in a medium that induces the regeneration of shoots in the plant species being transformed as described by Fraley *et al.*, *Proc. Nat'l. Acad. Sci. (U.S.A.)*, 80:4803 (1983). This procedure typically produces shoots within two to four weeks and these transformant shoots are then transferred to an appropriate root-inducing medium

containing the selective agent and an antibiotic to prevent bacterial growth. Transgenic plants of the present invention may be fertile or sterile.

Regeneration can also be obtained from plant callus, explants, organs, or parts thereof. Such regeneration techniques are described generally in Klee *et al.*, *Ann. Rev. of Plant Phys.* 38: 467-486 (1987). The regeneration of plants from either single plant protoplasts or various explants is well known in the art. See, for example, *Methods for Plant Molecular Biology*, A. Weissbach and H. Weissbach, eds., Academic Press, Inc., San Diego, Calif. (1988). This regeneration and growth process includes the steps of selection of transformant cells and shoots, rooting the transformant shoots and growth of the plantlets in soil. For maize cell culture and regeneration see generally, *The Maize Handbook*, Freeling and Walbot, Eds., Springer, New York (1994); *Corn and Corn Improvement*, 3rd edition, Sprague and Dudley Eds., American Society of Agronomy, Madison, Wisconsin (1988).

One of skill will recognize that after the recombinant expression cassette is stably incorporated in transgenic plants and confirmed to be operable, it can be introduced into other plants by sexual crossing. Any of a number of standard breeding techniques can be used, depending upon the species to be crossed.

In vegetatively propagated crops, mature transgenic plants can be propagated by the taking of cuttings or by tissue culture techniques to produce multiple identical plants. Selection of desirable transgenics is made and new varieties are obtained and propagated vegetatively for commercial use. In seed propagated crops, mature transgenic plants can be self crossed to produce a homozygous inbred plant. The inbred plant produces seed containing the newly introduced heterologous nucleic acid. These seeds can be grown to produce plants that would produce the selected phenotype.

Parts obtained from the regenerated plant, such as flowers, seeds, leaves, branches, fruit, and the like are included in the invention, provided that these parts comprise cells comprising the isolated nucleic acid of the present invention. Progeny and variants, and mutants of the regenerated plants are also included within the scope of the invention, provided that these parts comprise the introduced nucleic acid sequences.

Transgenic plants expressing the selectable marker can be screened for transmission of the nucleic acid of the present invention by, for example, standard immunoblot and DNA detection techniques. Transgenic lines are also typically evaluated on levels of expression of the heterologous nucleic acid. Expression at the RNA level can

be determined initially to identify and quantitate expression-positive plants. Standard techniques for RNA analysis can be employed and include PCR amplification assays using oligonucleotide primers designed to amplify only the heterologous RNA templates and solution hybridization assays using heterologous nucleic acid-specific probes. The RNA-positive plants can then analyzed for protein expression by Western immunoblot analysis using the specifically reactive antibodies of the present invention. In addition, *in situ* hybridization and immunocytochemistry according to standard protocols can be done using heterologous nucleic acid specific polynucleotide probes and antibodies, respectively, to localize sites of expression within transgenic tissue. Generally, a number of transgenic lines are usually screened for the incorporated nucleic acid to identify and select plants with the most appropriate expression profiles.

A preferred embodiment is a transgenic plant that is homozygous for the added heterologous nucleic acid; i.e., a transgenic plant that contains two added nucleic acid sequences, one gene at the same locus on each chromosome of a chromosome pair. A homozygous transgenic plant can be obtained by sexually mating (selfing) a heterozygous transgenic plant that contains a single added heterologous nucleic acid, germinating some of the seed produced and analyzing the resulting plants produced for altered expression of a polynucleotide of the present invention relative to a control plant (i.e., native, non-transgenic). Back-crossing to a parental plant and out-crossing with a non- transgenic plant are also contemplated.

Modulating Polypeptide Levels and/or Composition

The present invention further provides a method for modulating (i.e., increasing or decreasing) the concentration or composition of the polypeptides of the present invention in a plant or part thereof. Modulation can be effected by increasing or decreasing the concentration and/or the composition (i.e., the ratio of the polypeptides of the present invention) in a plant. The method comprises introducing into a plant cell a recombinant expression cassette comprising a polynucleotide of the present invention as described above to obtain a transformed plant cell, culturing the transformed plant cell under plant cell growing conditions, and inducing or repressing expression of a polynucleotide of the present invention in the plant for a time sufficient to modulate concentration and/or composition in the plant or plant part.

In some embodiments, the content and/or composition of polypeptides of the present invention in a plant may be modulated by altering, *in vivo* or *in vitro*, the promoter of a gene to up- or down-regulate gene expression. In some embodiments, the coding regions of native genes of the present invention can be altered via substitution, addition, insertion, or deletion to decrease activity of the encoded enzyme. See, e.g., Kmiec, U.S. Patent 5,565,350; Zarling *et al.*, WO 93/22443. And in some embodiments, an isolated nucleic acid (e.g., a vector) comprising a promoter sequence is transfected into a plant cell. Subsequently, a plant cell comprising the promoter operably linked to a polynucleotide of the present invention is selected for by means known to those of skill in the art such as, but not limited to, Southern blot, DNA sequencing, or PCR analysis using primers specific to the promoter and to the gene and detecting amplicons produced therefrom. A plant or plant part altered or modified by the foregoing embodiments is grown under plant forming conditions for a time sufficient to modulate the concentration and/or composition of polypeptides of the present invention in the plant. Plant forming conditions are well known in the art and discussed briefly, *supra*.

In general, concentration or composition is increased or decreased by at least 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90% relative to a native control plant, plant part, or cell lacking the aforementioned recombinant expression cassette. Modulation in the present invention may occur during and/or subsequent to growth of the plant to the desired stage of development. Modulating nucleic acid expression temporally and/or in particular tissues can be controlled by employing the appropriate promoter operably linked to a polynucleotide of the present invention in, for example, sense or antisense orientation as discussed in greater detail, *supra*. Induction of expression of a polynucleotide of the present invention can also be controlled by exogenous administration of an effective amount of inducing compound. Inducible promoters and inducing compounds which activate expression from these promoters are well known in the art. In preferred embodiments, the polypeptides of the present invention are modulated in monocots, particularly maize.

Molecular Markers

The present invention provides a method of genotyping a plant comprising a polynucleotide of the present invention. Optionally, the plant is a monocot, such as maize or sorghum. Genotyping provides a means of distinguishing homologs of a chromosome

pair and can be used to differentiate segregants in a plant population. Molecular marker methods can be used for phylogenetic studies, characterizing genetic relationships among crop varieties, identifying crosses or somatic hybrids, localizing chromosomal segments affecting monogenic traits, map based cloning, and the study of quantitative inheritance. See, e.g., Clark, Ed., *Plant Molecular Biology: A Laboratory Manual.* Berlin, Springer-Verlag, 1997. Chapter 7. For molecular marker methods, see generally, "The DNA Revolution" in: Paterson, A.H., *Genome Mapping in Plants* (Austin, TX, Academic Press/R. G. Landis Company, 1996) pp.7-21.

The particular method of genotyping in the present invention may employ any number of molecular marker analytic techniques such as, but not limited to, restriction fragment length polymorphisms (RFLPs). RFLPs are the product of allelic differences between DNA restriction fragments resulting from nucleotide sequence variability. As is well known to those of skill in the art, RFLPs are typically detected by extraction of genomic DNA and digestion with a restriction enzyme. Generally, the resulting fragments are separated according to size and hybridized with a probe; single copy probes are preferred. Restriction fragments from homologous chromosomes are revealed. Differences in fragment size among alleles represent an RFLP. Thus, the present invention further provides a means to follow segregation of a gene or nucleic acid of the present invention as well as chromosomal sequences genetically linked to these genes or nucleic acids using such techniques as RFLP analysis. Linked chromosomal sequences are within 50 centiMorgans (cM), often within 40 or 30 cM, preferably within 20 or 10 cM, more preferably within 5, 3, 2, or 1 cM of a gene of the present invention.

In the present invention, the nucleic acid probes employed for molecular marker mapping of plant nuclear genomes selectively hybridize, under selective hybridization conditions, to a gene encoding a polynucleotide of the present invention. In preferred embodiments, the probes are selected from polynucleotides of the present invention. Typically, these probes are cDNA probes or restriction-enzyme treated (e.g., *Pst I)* genomic clones. The length of the probes is discussed in greater detail, *supra*, but are typically at least 15 bases in length, more preferably at least 20, 25, 30, 35, 40, or 50 bases in length. Generally, however, the probes are less than about 1 kilobase in length. Preferably, the probes are single copy probes that hybridize to a unique locus in a haploid chromosome complement. Some exemplary restriction enzymes employed in RFLP mapping are *Eco*RI, *Eco*Rv, and *Sst*I. As used herein the term "restriction enzyme"

includes reference to a composition that recognizes and, alone or in conjunction with another composition, cleaves at a specific nucleotide sequence.

The method of detecting an RFLP comprises the steps of (a) digesting genomic DNA of a plant with a restriction enzyme; (b) hybridizing a nucleic acid probe, under selective hybridization conditions, to a sequence of a polynucleotide of the present of said genomic DNA; (c) detecting therefrom a RFLP. Other methods of differentiating polymorphic (allelic) variants of polynucleotides of the present invention can be had by utilizing molecular marker techniques well known to those of skill in the art including such techniques as: 1) single stranded conformation analysis (SSCA); 2) denaturing gradient gel electrophoresis (DGGE); 3) RNase protection assays; 4) allele-specific oligonucleotides (ASOs); 5) the use of proteins which recognize nucleotide mismatches, such as the *E. coli* mutS protein; and 6) allele-specific PCR. Other approaches based on the detection of mismatches between the two complementary DNA strands include clamped denaturing gel electrophoresis (CDGE); heteroduplex analysis (HA); and chemical mismatch cleavage (CMC). Thus, the present invention further provides a method of genotyping comprising the steps of contacting, under stringent hybridization conditions, a sample suspected of comprising a polynucleotide of the present invention with a nucleic acid probe. Generally, the sample is a plant sample; preferably, a sample suspected of comprising a maize polynucleotide of the present invention (e.g., gene, mRNA). The nucleic acid probe selectively hybridizes, under stringent conditions, to a subsequence of a polynucleotide of the present invention comprising a polymorphic marker. Selective hybridization of the nucleic acid probe to the polymorphic marker nucleic acid sequence yields a hybridization complex. Detection of the hybridization complex indicates the presence of that polymorphic marker in the sample. In preferred embodiments, the nucleic acid probe comprises a polynucleotide of the present invention.

UTRs and Codon Preference

In general, translational efficiency has been found to be regulated by specific sequence elements in the 5' non-coding or untranslated region (5' UTR) of the RNA. Positive sequence motifs include translational initiation consensus sequences (Kozak, *Nucleic Acids Res.* 15:8125 (1987)) and the 7-methylguanosine cap structure (Drummond *et al., Nucleic Acids Res.* 13:7375 (1985)). Negative elements include stable intramolecular 5' UTR stem-loop structures (Muesing *et al., Cell* 48:691 (1987)) and AUG

sequences or short open reading frames preceded by an appropriate AUG in the 5' UTR (Kozak, *supra*, Rao *et al.*, *Mol. and Cell. Biol.* 8:284 (1988)). Accordingly, the present invention provides 5' and/or 3' untranslated regions for modulation of translation of heterologous coding sequences.

Further, the polypeptide-encoding segments of the polynucleotides of the present invention can be modified to alter codon usage. Altered codon usage can be employed to alter translational efficiency and/or to optimize the coding sequence for expression in a desired host such as to optimize the codon usage in a heterologous sequence for expression in maize. Codon usage in the coding regions of the polynucleotides of the present invention can be analyzed statistically using commercially available software packages such as "Codon Preference" available from the University of Wisconsin Genetics Computer Group (see Devereaux *et al.*, *Nucleic Acids Res.* 12: 387-395 (1984)) or MacVector 4.1 (Eastman Kodak Co., New Haven, Conn.). Thus, the present invention provides a codon usage frequency characteristic of the coding region of at least one of the polynucleotides of the present invention. The number of polynucleotides that can be used to determine a codon usage frequency can be any integer from 1 to the number of polynucleotides of the present invention as provided herein. Optionally, the polynucleotides will be full-length sequences. An exemplary number of sequences for statistical analysis can be at least 1, 5, 10, 20, 50, or 100.

Sequence Shuffling

The present invention provides methods for sequence shuffling using polynucleotides of the present invention, and compositions resulting therefrom. Sequence shuffling is described in PCT publication No. WO 96/19256. See also, Zhang, J.- H., *et al.* *Proc. Nat'l. Acad. Sci. USA* 94:4504-4509 (1997). Generally, sequence shuffling provides a means for generating libraries of polynucleotides having a desired characteristic which can be selected or screened for. Libraries of recombinant polynucleotides are generated from a population of related sequence polynucleotides which comprise sequence regions which have substantial sequence identity and can be homologously recombined in vitro or in vivo. The population of sequence-recombined polynucleotides comprises a subpopulation of polynucleotides which possess desired or advantageous characteristics and which can be selected by a suitable selection or screening method. The characteristics can be any property or attribute capable of being selected for or detected in a screening

system, and may include properties of: an encoded protein, a transcriptional element, a sequence controlling transcription, RNA processing, RNA stability, chromatin conformation, translation, or other expression property of a gene or transgene, a replicative element, a protein-binding element, or the like, such as any feature which confers a selectable or detectable property. In some embodiments, the selected characteristic will be a decreased K_m and/or increased K_{cat} over the wild-type protein as provided herein. In other embodiments, a protein or polynucleotide generated from sequence shuffling will have a ligand binding affinity greater than the non-shuffled wild-type polynucleotide. The increase in such properties can be at least 110%, 120%, 130%, 140% or at least 150% of the wild-type value.

Generic and Consensus Sequences

Polynucleotides and polypeptides of the present invention further include those having: (a) a generic sequence of at least two homologous polynucleotides or polypeptides, respectively, of the present invention; and, (b) a consensus sequence of at least three homologous polynucleotides or polypeptides, respectively, of the present invention. The generic sequence of the present invention comprises each species of polypeptide or polynucleotide embraced by the generic polypeptide or polynucleotide sequence, respectively. The individual species encompassed by a polynucleotide having an amino acid or nucleic acid consensus sequence can be used to generate antibodies or produce nucleic acid probes or primers to screen for homologs in other species, genera, families, orders, classes, phyla, or kingdoms. For example, a polynucleotide having a consensus sequence from a gene family of *Zea mays* can be used to generate antibody or nucleic acid probes or primers to other *Gramineae* species such as wheat, rice, or sorghum. Alternatively, a polynucleotide having a consensus sequence generated from orthologous genes can be used to identify or isolate orthologs of other taxa. Typically, a polynucleotide having a consensus sequence will be at least 9, 10, 15, 20, 25, 30, or 40 amino acids in length, or 20, 30, 40, 50, 100, or 150 nucleotides in length. As those of skill in the art are aware, a conservative amino acid substitution can be used for amino acids which differ amongst aligned sequence but are from the same conservative substitution group as discussed above. Optionally, no more than 1 or 2 conservative amino acids are substituted for each 10 amino acid length of consensus sequence.

Similar sequences used for generation of a consensus or generic sequence include any number and combination of allelic variants of the same gene, orthologous, or paralogous sequences as provided herein. Optionally, similar sequences used in generating a consensus or generic sequence are identified using the BLAST algorithm's smallest sum probability (P(N)). Various suppliers of sequence-analysis software are listed in chapter 7 of *Current Protocols in Molecular Biology*, F.M. Ausubel *et al.*, Eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc. (Supplement 30). A polynucleotide sequence is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, more preferably less than about 0.01, or 0.001, and most preferably less than about 0.0001, or 0.00001. Similar polynucleotides can be aligned and a consensus or generic sequence generated using multiple sequence alignment software available from a number of commercial suppliers such as the Genetics Computer Group's (Madison, WI) PILEUP software, Vector NTI's (North Bethesda, MD) ALIGNX, or Genecode's (Ann Arbor, MI) SEQUENCHER. Conveniently, default parameters of such software can be used to generate consensus or generic sequences.

Assays for Compounds that Modulate Enzymatic Activity or Expression

The present invention also provides means for identifying compounds that bind to (e.g., substrates), and/or increase or decrease (i.e., modulate) the enzymatic activity of, catalytically active polypeptides of the present invention. The method comprises contacting a polypeptide of the present invention with a compound whose ability to bind to or modulate enzyme activity is to be determined. The polypeptide employed will have at least 20%, preferably at least 30% or 40%, more preferably at least 50% or 60%, and most preferably at least 70% or 80% of the specific activity of the native, full-length polypeptide of the present invention (e.g., enzyme). Generally, the polypeptide will be present in a range sufficient to determine the effect of the compound, typically about 1 nM to 10 μM. Likewise, the compound will be present in a concentration of from about 1 nM to 10 μM. Those of skill will understand that such factors as enzyme concentration, ligand concentrations (i.e., substrates, products, inhibitors, activators), pH, ionic strength, and temperature will be controlled so as to obtain useful kinetic data and determine the presence of absence of a compound that binds or modulates polypeptide activity. Methods

of measuring enzyme kinetics is well known in the art. See, e.g., Segel, *Biochemical Calculations*, 2nd ed., John Wiley and Sons, New York (1976).

Although the present invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be obvious that certain changes and modifications may be practiced within the scope of the appended claims.

Examples

The present invention is further described by the following examples. The examples are provided solely to illustrate the invention by reference to specific embodiments. These exemplifications, while illustrating certain specific aspects of the invention, do not portray the limitations or circumscribe the scope of the disclosed invention. It will be obvious that certain changes and modifications may be practiced within the scope of the appended claims.

Example 1

This example describes the construction of a cDNA library.

Total RNA can be isolated from maize tissues with TRIzol Reagent (Life Technology Inc. Gaithersburg, MD) using a modification of the guanidine isothiocyanate/acid-phenol procedure described by Chomczynski and Sacchi (Chomczynski, P., and Sacchi, N. *Anal. Biochem.* 162, 156 (1987)). In brief, plant tissue samples are pulverized in liquid nitrogen before the addition of the TRIzol Reagent, and then further homogenized with a mortar and pestle. Addition of chloroform followed by centrifugation is conducted for separation of an aqueous phase and an organic phase. The total RNA is recovered by precipitation with isopropyl alcohol from the aqueous phase.

The selection of poly(A)+ RNA from total RNA can be performed using PolyATact system (Promega Corporation. Madison, WI). Biotinylated oligo(dT) primers are used to hybridize to the 3' poly(A) tails on mRNA. The hybrids are captured using streptavidin coupled to paramagnetic particles and a magnetic separation stand. The mRNA is then washed at high stringency conditions and eluted by RNase-free deionized water.

cDNA synthesis and construction of unidirectional cDNA libraries can be accomplished using the SuperScript Plasmid System (Life Technology Inc. Gaithersburg, MD). The first strand of cDNA is synthesized by priming an oligo(dT) primer containing a

Not I site. The reaction is catalyzed by SuperScript Reverse Transcriptase II at 45°C. The second strand of cDNA is labeled with alpha-^{32}P-dCTP and a portion of the reaction analyzed by agarose gel electrophoresis to determine cDNA sizes. cDNA molecules smaller than 500 base pairs and unligated adapters are removed by Sephacryl-S400 chromatography. The selected cDNA molecules are ligated into pSPORT1 vector in between of *Not* I and *Sal* I sites.

Alternatively, cDNA libraries can be prepared by any one of many methods available. For example, the cDNAs may be introduced into plasmid vectors by first preparing the cDNA libraries in Uni-ZAP™ XR vectors according to the manufacturer's protocol (Stratagene Cloning Systems, La Jolla, CA). The Uni-ZAP™ XR libraries are converted into plasmid libraries according to the protocol provided by Stratagene. Upon conversion, cDNA inserts will be contained in the plasmid vector pBluescript. In addition, the cDNAs may be introduced directly into precut Bluescript II SK(+) vectors (Stratagene) using T4 DNA ligase (New England Biolabs), followed by transfection into DH10B cells according to the manufacturer's protocol (GIBCO BRL Products). Once the cDNA inserts are in plasmid vectors, plasmid DNAs are prepared from randomly picked bacterial colonies containing recombinant pBluescript plasmids, or the insert cDNA sequences are amplified via polymerase chain reaction using primers specific for vector sequences flanking the inserted cDNA sequences. Amplified insert DNAs or plasmid DNAs are sequenced in dye-primer sequencing reactions to generate partial cDNA sequences (expressed sequence tags or "ESTs"; see Adams *et al.*, (1991) *Science 252*:1651-1656). The resulting ESTs are analyzed using a Perkin Elmer Model 377 fluorescent sequencer.

Example 2

This method describes construction of a full-length enriched cDNA library.

An enriched full-length cDNA library can be constructed using one of two variations of the method of Carninci *et al. Genomics* 37: 327-336, 1996. These variations are based on chemical introduction of a biotin group into the diol residue of the 5' cap structure of eukaryotic mRNA to select full-length first strand cDNA. The selection occurs by trapping the biotin residue at the cap sites using streptavidin-coated magnetic beads followed by RNase I treatment to eliminate incompletely synthesized cDNAs. Second strand cDNA is synthesized using established procedures such as those provided in Life Technologies' (Rockville, MD) "SuperScript Plasmid System for cDNA Synthesis

and Plasmid Cloning" kit. Libraries made by this method have been shown to contain 50% to 70% full-length cDNAs.

The first strand synthesis methods are detailed below. An asterisk denotes that the reagent was obtained from Life Technologies, Inc.

A. First strand cDNA synthesis method 1 (with trehalose)

mRNA (10ug)	25µl
*Not I primer (5ug)	10µl
*5x 1st strand buffer	43µl
*0.1m DTT	20µl
*dNTP mix 10mm	10µl
BSA 10ug/µl	1µl
Trehalose (saturated)	59.2µl
RNase inhibitor (Promega)	1.8µl
*Superscript II RT 200u/µl	20µl
100 % glycerol	18µl
Water	7µl

The mRNA and Not I primer are mixed and denatured at 65°C for 10 min. They are then chilled on ice and other components added to the tube. Incubation is at 45°C for 2 min. Twenty microliters of RT (reverse transcriptase) is added to the reaction and start program on the thermocycler (MJ Research, Waltham, MA):

Step 1	45°C 10min
Step 2	45°C –0.3°C/cycle , 2 seconds/cycle
Step 3	go to 2 for 33 cycles
Step 4	35°C 5min
Step 5	45°C 5min
Step 6	45°C 0.2°C/cycle, 1 sec/cycle
Step 7	go to 7 for 49 cycles
Step 8	55°C 0.1°C/cycle, 12 sec/cycle
Step 9	go to 8 for 49 cycles
Step 10	55°C 2min

Step11	60°C 2min
Step 12	go to 11 for 9 times
Step 13	4°C forever
Step14	end

B. First strand cDNA synthesis method 2

mRNA (10μg)	25μl
water	30μl
*Not I adapter primer (5μg)	10μl

65°C for 10min, chill on ice, then add following reagents,

*5x first buffer	20μl
*0.1M DTT	10μl
*10mM dNTP mix	5μl

Incubate at 45°C for 2min, then add 10μl of *Superscript II RT (200u/μl), start the following program:

Step 1	45°C for 6 sec, -0.1°C/cycle
Step 2	go to 1 for 99 additional cycles
Step 3	35°C for 5min
Step 4	45°C for 60 min
Step 5	50°C for 10 min
Step 6	4°C forever
Step 7	end

After the 1st strand cDNA synthesis, the DNA is extracted by phenol according to standard procedures, and then precipitated in NaOAc and ethanol, and stored in –20°C.

C. Oxidization of the diol group of mRNA for biotin labeling

First strand cDNA is spun down and washed once with 70% EtOH. The pellet resuspended in 23.2 μl of DEPC treated water and put on ice. Prepare 100 mM of NaIO4 freshly, and then add the following reagents:

mRNA:1st cDNA (start with 20µg mRNA)	46.4µl
100mM NaIO4 (freshly made)	2.5µl
NaOAc 3M pH4.5	1.1µl

To make 100 mM NaIO4, use 21.39µg of NaIO4 for 1µl of water.

Wrap the tube in a foil and incubate on ice for 45min.

After the incubation, the reaction is then precipitated in:

5M NaCl	10µl
20%SDS	0.5µl
isopropanol	61µl

Incubate on ice for at least 30 min, then spin it down at max speed at 4°C for 30 min and wash once with 70% ethanol and then 80% EtOH.

D. Biotinylation of the mRNA diol group

Resuspend the DNA in 110µl DEPC treated water, then add the following reagents:

20% SDS	5 µl
2 M NaOAc pH 6.1	5 µl
10mm biotin hydrazide (freshly made)	300 µl

Wrap in a foil and incubate at room temperature overnight.

E. RNase I treatment

Precipitate DNA in:

5M NaCl	10µl
2M NaOAc pH 6.1	75µl
biotinylated mRNA:cDNA	420µl
100% EtOH (2.5Vol)	1262.5µl

(Perform this precipitation in two tubes and split the 420 µl of DNA into 210 µl each, add 5µl of 5M NaCl, 37.5µl of 2M NaOAc pH 6.1, and 631.25 µl of 100% EtOH).

Store at –20°C for at least 30 min. Spin the DNA down at 4°C at maximal speed for 30 min. and wash with 80% EtOH twice, then dissolve DNA in 70µl RNase free water. Pool two tubes and end up with 140 µl.

Add the following reagents:

RNase One 10U/µl	40µl
1st cDNA:RNA	140µl
10X buffer	20µl

Incubate at 37°C for 15min.

Add 5µl of 40µg/µl yeast tRNA to each sample for capturing.

F. Full length 1st cDNA capturing

Blocking the beads with yeast tRNA:

Beads	1ml
Yeast tRNA 40µg/µl	5µl

Incubate on ice for 30min with mixing, wash 3 times with 1ml of 2M NaCl , 50mmEDTA, pH 8.0.

Resuspend the beads in 800µl of 2M NaCl , 50mm EDTA, pH 8.0, add RNase I treated sample 200µl, and incubate the reaction for 30min at room temperature. Capture the beads using the magnetic stand, save the supernatant, and start following washes:

2 washes with 2M NaCl , 50mm EDTA, pH 8.0, 1 ml each time,
1 wash with 0.4% SDS, 50µg/ml tRNA,
1 wash with 10mm Tris-Cl pH 7.5, 0.2mm EDTA, 10mm NaCl, 20% glycerol,
1 wash with 50µg/ml tRNA,
1 wash with 1st cDNA buffer

G. Second strand cDNA synthesis

Resuspend the beads in:

*5X first buffer	8µl
*0.1mM DTT	4µl
*10mm dNTP mix	8µl
*5X 2nd buffer	60µl
*E.coli Ligase 10U/µl	2µl

*E.coli DNA polymerase 10U/μl	8μl
*E. coli RNaseH 2U/μl	2μl
P32 dCTP 10μci/μl	2μl
Or water up to 300μl	208μl

Incubate at 16°C for 2hr with mixing the reaction in every 30 min.

Add 4μl of T4 DNA polymerase and incubate for additional 5 min at 16°C.

Elute 2nd cDNA from the beads.

Use a magnetic stand to separate the 2nd cDNA from the beads, then resuspend the beads in 200μl of water, and then separate again, pool the samples (about 500μl),

Add 200 μl of water to the beads, then 200μl of phenol:chloroform, vortex, and spin to separate the sample with phenol.

Pool the DNA together (about 700μl) and use phenol to clean the DNA again, DNA is then precipitated in 2μg of glycogen and 0.5 vol of 7.5M NH4OAc and 2 vol of 100% EtOH.

Precipitate overnight. Spin down the pellet and wash with 70% EtOH, air-dry the pellet.

DNA	250μl	DNA	200μl
7.5M NH4OAc	125μl	7.5M NH4OAc	100μl
100% EtOH	750μl	100% EtOH	600μl
glycogen 1μg/μl	2μl	glycogen 1μg/μl	2μl

H. Sal I adapter ligation

Resuspend the pellet in 26 μl of water and use 1μl for TAE gel.

Set up reaction as following:

2nd strand cDNA	25μl
*5X T4 DNA ligase buffer	10μl
*Sal I adapters	10μl
*T4 DNA ligase	5μl

Mix gently, incubate the reaction at 16°C overnight.

Add 2μl of ligase second day and incubate at room temperature for 2 hrs (optional).

Add 50µl water to the reaction and use 100µl of phenol to clean the DNA, 90µl of the upper phase is transferred into a new tube and precipitate in:

Glycogen 1µg/µl	2µl
Upper phase DNA	90µl
7.5M NH4OAc	50µl
100% EtOH	300µl

precipitate at –20°C overnight

Spin down the pellet at 4°C and wash in 70% EtOH, dry the pellet.

I. Not I digestion

2nd cDNA	41µl
*Reaction 3 buffer	5µl
*Not I 15u/µl	4µl

Mix gently and incubate the reaction at 37°C for 2hr.

Add 50 µl of water and 100µl of phenol, vortex , and take 90µl of the upper phase to a new tube, then add 50µl of NH40Ac and 300 µl of EtOH. Precipitate overnight at -20°C.

Cloning, ligation, and transformation are performed per the Superscript cDNA synthesis kit.

Example 3

This example describes cDNA sequencing and library subtraction.

Individual colonies can be picked and DNA prepared either by PCR with M13 forward primers and M13 reverse primers, or by plasmid isolation. cDNA clones can be sequenced using M13 reverse primers.

cDNA libraries are plated out on 22 x 22 cm^2 agar plate at density of about 3,000 colonies per plate. The plates are incubated in a 37°C incubator for 12-24 hours. Colonies are picked into 384-well plates by a robot colony picker, Q-bot (GENETIX Limited). These plates are incubated overnight at 37°C. Once sufficient colonies are picked, they are pinned onto 22 x 22 cm^2 nylon membranes using Q-bot. Each membrane holds 9,216 or 36,864 colonies. These membranes are placed onto an agar plate with an appropriate antibiotic. The plates are incubated at 37°C overnight.

After colonies are recovered on the second day, these filters are placed on filter paper prewetted with denaturing solution for four minutes, then incubated on top of a boiling water bath for an additional four minutes. The filters are then placed on filter paper prewetted with neutralizing solution for four minutes. After excess solution is removed by placing the filters on dry filter papers for one minute, the colony side of the filters is placed into Proteinase K solution, incubated at 37°C for 40-50 minutes. The filters are placed on dry filter papers to dry overnight. DNA is then cross-linked to nylon membrane by UV light treatment.

Colony hybridization is conducted as described by Sambrook,J., Fritsch, E.F. and Maniatis, T., (in Molecular Cloning: A laboratory Manual, 2nd Edition). The following probes can be used in colony hybridization:

1. First strand cDNA from the same tissue as the library was made from to remove the most redundant clones.

2. 48-192 most redundant cDNA clones from the same library based on previous sequencing data.

3. 192 most redundant cDNA clones in the entire maize sequence database.

4. A Sal-A20 oligo nucleotide: TCG ACC CAC GCG TCC GAA AAA AAA AAA AAA AAA AAA AAA, removes clones containing a poly A tail but no cDNA.

5. cDNA clones derived from rRNA.

The image of the autoradiography is scanned into computer and the signal intensity and cold colony addresses of each colony is analyzed. Re-arraying of cold-colonies from 384 well plates to 96 well plates is conducted using Q-bot.

Example 4

This example describes identification of the gene from a computer homology search.

Gene identities can be determined by conducting BLAST (Basic Local Alignment Search Tool; Altschul, S. F., et al., (1993) J. Mol. Biol. 215:403-410; see also www.ncbi.nlm.nih.gov/BLAST/) searches under default parameters for similarity to sequences contained in the BLAST "nr" database (comprising all non-redundant GenBank CDS translations, sequences derived from the 3-dimensional structure Brookhaven Protein Data Bank, the last major release of the SWISS-PROT protein sequence database, EMBL, and DDBJ databases). The cDNA sequences are analyzed for similarity to all publicly

available DNA sequences contained in the "nr" database using the BLASTN algorithm. The DNA sequences are translated in all reading frames and compared for similarity to all publicly available protein sequences contained in the "nr" database using the BLASTX algorithm (Gish, W. and States, D. J. *Nature Genetics 3*:266-272 (1993)) provided by the NCBI. In some cases, the sequencing data from two or more clones containing overlapping segments of DNA are used to construct contiguous DNA sequences.

Sequence alignments and percent identity calculations can be performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences can be performed using the Clustal method of alignment (Higgins and Sharp (1989) *CABIOS. 5*:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method are KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5.

Example 5

This example describes expression of transgenes in monocot cells.

A transgene comprising a cDNA encoding the instant polypeptides in sense orientation with respect to the maize 27 kD zein promoter that is located 5' to the cDNA fragment, and the 10 kD zein 3' end that is located 3' to the cDNA fragment, can be constructed. The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites (NcoI or SmaI) can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the digested vector pML103 as described below. Amplification is then performed in a standard PCR. The amplified DNA is then digested with restriction enzymes NcoI and SmaI and fractionated on an agarose gel. The appropriate band can be isolated from the gel and combined with a 4.9 kb NcoI-SmaI fragment of the plasmid pML103. Plasmid pML103 has been deposited under the terms of the Budapest Treaty at ATCC (American Type Culture Collection, 10801 University Blvd., Manassas, VA 20110-2209), and bears accession number ATCC 97366. The DNA segment from pML103 contains a 1.05 kb SalI-NcoI promoter fragment of the maize 27 kD zein gene and a 0.96 kb SmaI-SalI fragment from the 3' end of the maize 10 kD zein gene in the vector pGem9Zf(+) (Promega). Vector and insert DNA can be ligated at 15°C overnight, essentially as described (Maniatis). The ligated DNA may then

be used to transform *E. coli* XL1-Blue (Epicurian Coli XL-1 Blue; Stratagene). Bacterial transformants can be screened by restriction enzyme digestion of plasmid DNA and limited nucleotide sequence analysis using the dideoxy chain termination method (Sequenase DNA Sequencing Kit; U. S. Biochemical). The resulting plasmid construct would comprise a transgene encoding, in the 5' to 3' direction, the maize 27 kD zein promoter, a cDNA fragment encoding the instant polypeptides, and the 10 kD zein 3' region.

The transgene described above can then be introduced into maize cells by the following procedure. Immature maize embryos can be dissected from developing caryopses derived from crosses of the inbred maize lines H99 and LH132. The embryos are isolated 10 to 11 days after pollination when they are 1.0 to 1.5 mm long. The embryos are then placed with the axis-side facing down and in contact with agarose-solidified N6 medium (Chu *et al.* (1975) *Sci. Sin. Peking* 18:659-668). The embryos are kept in the dark at 27°C. Friable embryogenic callus consisting of undifferentiated masses of cells with somatic proembryoids and embryoids borne on suspensor structures proliferates from the scutellum of these immature embryos. The embryogenic callus isolated from the primary explant can be cultured on N6 medium and sub-cultured on this medium every 2 to 3 weeks.

The plasmid, p35S/Ac (Hoechst Ag, Frankfurt, Germany) or equivalent may be used in transformation experiments in order to provide for a selectable marker. This plasmid contains the *Pat* gene (see European Patent Publication 0 242 236) which encodes phosphinothricin acetyl transferase (PAT). The enzyme PAT confers resistance to herbicidal glutamine synthetase inhibitors such as phosphinothricin. The *pat* gene in p35S/Ac is under the control of the 35S promoter from Cauliflower Mosaic Virus (Odell et al. (1985) *Nature* 313:810-812) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*.

The particle bombardment method (Klein *et al.* (1987) *Nature* 327:70-73) may be used to transfer genes to the callus culture cells. According to this method, gold particles (1 μm in diameter) are coated with DNA using the following technique. Ten μg of plasmid DNAs are added to 50 μL of a suspension of gold particles (60 mg per mL). Calcium chloride (50 μL of a 2.5 M solution) and spermidine free base (20 μL of a 1.0 M solution) are added to the particles. The suspension is vortexed during the addition of these solutions. After 10 minutes, the tubes are briefly centrifuged (5 sec at 15,000 rpm)

and the supernatant removed. The particles are resuspended in 200 µL of absolute ethanol, centrifuged again and the supernatant removed. The ethanol rinse is performed again and the particles resuspended in a final volume of 30 µL of ethanol. An aliquot (5 µL) of the DNA-coated gold particles can be placed in the center of a Kapton flying disc (Bio-Rad Labs). The particles are then accelerated into the maize tissue with a Biolistic PDS-1000/He (Bio-Rad Instruments, Hercules CA), using a helium pressure of 1000 psi, a gap distance of 0.5 cm and a flying distance of 1.0 cm.

For bombardment, the embryogenic tissue is placed on filter paper over agarose-solidified N6 medium. The tissue is arranged as a thin lawn and covers a circular area of about 5 cm in diameter. The petri dish containing the tissue can be placed in the chamber of the PDS-1000/He approximately 8 cm from the stopping screen. The air in the chamber is then evacuated to a vacuum of 28 inches of Hg. The macrocarrier is accelerated with a helium shock wave using a rupture membrane that bursts when the He pressure in the shock tube reaches 1000 psi.

Seven days after bombardment the tissue can be transferred to N6 medium that contains gluphosinate (2 mg per liter) and lacks casein or proline. The tissue continues to grow slowly on this medium. After an additional 2 weeks the tissue can be transferred to fresh N6 medium containing gluphosinate. After 6 weeks, areas of about 1 cm in diameter of actively growing callus can be identified on some of the plates containing the glufosinate-supplemented medium. These calli may continue to grow when sub-cultured on the selective medium.

Plants can be regenerated from the transgenic callus by first transferring clusters of tissue to N6 medium supplemented with 0.2 mg per liter of 2,4-D. After two weeks the tissue can be transferred to regeneration medium (Fromm *et al.* (1990) *Bio/Technology* *8*:833-839).

Example 6

This example describes expression of transgenes in dicot cells.

A seed-specific expression cassette composed of the promoter and transcription terminator from the gene encoding the β subunit of the seed storage protein phaseolin from the bean *Phaseolus vulgaris* (Doyle *et al.* (1986) *J. Biol. Chem.* *261*:9228-9238) can be used for expression of the instant polypeptides in transformed soybean. The phaseolin cassette includes about 500 nucleotides upstream (5') from the translation initiation codon

and about 1650 nucleotides downstream (3') from the translation stop codon of phaseolin. Between the 5' and 3' regions are the unique restriction endonuclease sites Nco I (which includes the ATG translation initiation codon), SmaI, KpnI and XbaI. The entire cassette is flanked by Hind III sites.

The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the expression vector. Amplification is then performed as described above, and the isolated fragment is inserted into a pUC18 vector carrying the seed expression cassette.

Soybean embroys may then be transformed with the expression vector comprising sequences encoding the instant polypeptides. To induce somatic embryos, cotyledons, 3-5 mm in length dissected from surface sterilized, immature seeds of the soybean cultivar A2872, can be cultured in the light or dark at 26°C on an appropriate agar medium for 6-10 weeks. Somatic embryos which produce secondary embryos are then excised and placed into a suitable liquid medium. After repeated selection for clusters of somatic embryos which multiplied as early, globular staged embryos, the suspensions are maintained as described below.

Soybean embryogenic suspension cultures can be maintained in 35 mL liquid media on a rotary shaker, 150 rpm, at 26°C with florescent lights on a 16:8 hour day/night schedule. Cultures are subcultured every two weeks by inoculating approximately 35 mg of tissue into 35 mL of liquid medium.

Soybean embryogenic suspension cultures may then be transformed by the method of particle gun bombardment (Klein *et al.* (1987) *Nature* (London) *327*:70-73, U.S. Patent No. 4,945,050). A Du Pont Biolistic PDS1000/HE instrument (helium retrofit) can be used for these transformations.

A selectable marker gene which can be used to facilitate soybean transformation is a transgene composed of the 35S promoter from Cauliflower Mosaic Virus (Odell *et al.*(1985) *Nature 313*:810-812), the hygromycin phosphotransferase gene from plasmid pJR225 (from *E. coli*; Gritz *et al.*(1983) *Gene 25*:179-188) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*. The seed expression cassette comprising the phaseolin 5' region, the fragment encoding the instant polypeptides and the phaseolin 3' region can be isolated as a restriction fragment.

This fragment can then be inserted into a unique restriction site of the vector carrying the marker gene.

To 50 µL of a 60 mg/mL 1 m gold particle suspension is added (in order): 5 µL DNA (1 µg/µL), 20 µl spermidine (0.1 M), and 50 µL $CaCl_2$ (2.5 M). The particle preparation is then agitated for three minutes, spun in a microfuge for 10 seconds and the supernatant removed. The DNA-coated particles are then washed once in 400 µL 70% ethanol and resuspended in 40 µL of anhydrous ethanol. The DNA/particle suspension can be sonicated three times for one second each. Five microliters of the DNA-coated gold particles are then loaded on each macro carrier disk.

Approximately 300-400 mg of a two-week-old suspension culture is placed in an empty 60x15 mm petri dish and the residual liquid removed from the tissue with a pipette. For each transformation experiment, approximately 5-10 plates of tissue are normally bombarded. Membrane rupture pressure is set at 1100 psi and the chamber is evacuated to a vacuum of 28 inches mercury. The tissue is placed approximately 3.5 inches away from the retaining screen and bombarded three times. Following bombardment, the tissue can be divided in half and placed back into liquid and cultured as described above.

Five to seven days post bombardment, the liquid media may be exchanged with fresh media, and eleven to twelve days post bombardment with fresh media containing 50 mg/mL hygromycin. This selective media can be refreshed weekly. Seven to eight weeks post bombardment, green, transformed tissue may be observed growing from untransformed, necrotic embryogenic clusters. Isolated green tissue is removed and inoculated into individual flasks to generate new, clonally propagated, transformed embryogenic suspension cultures. Each new line may be treated as an independent transformation event. These suspensions can then be subcultured and maintained as clusters of immature embryos or regenerated into whole plants by maturation and germination of individual somatic embryos.

Example 7

This example describes expression of a transgene in microbial cells.

The cDNAs encoding the instant polypeptides can be inserted into the T7 *E. coli* expression vector pBT430. This vector is a derivative of pET-3a (Rosenberg *et al.* (1987) *Gene 56*:125-135) which employs the bacteriophage T7 RNA polymerase/T7 promoter system. Plasmid pBT430 was constructed by first destroying the EcoR I and Hind III sites

in pET-3a at their original positions. An oligonucleotide adaptor containing EcoR I and Hind III sites was inserted at the BamH I site of pET-3a. This created pET-3aM with additional unique cloning sites for insertion of genes into the expression vector. Then, the Nde I site at the position of translation initiation was converted to an Nco I site using oligonucleotide-directed mutagenesis. The DNA sequence of pET-3aM in this region, 5'-CATATGG, was converted to 5'-CCCATGG in pBT430.

Plasmid DNA containing a cDNA may be appropriately digested to release a nucleic acid fragment encoding the protein. This fragment may then be purified on a 1% NuSieve GTG low melting agarose gel (FMC). Buffer and agarose contain 10 μg/ml ethidium bromide for visualization of the DNA fragment. The fragment can then be purified from the agarose gel by digestion with GELase (Epicentre Technologies) according to the manufacturer's instructions, ethanol precipitated, dried and resuspended in 20 μL of water. Appropriate oligonucleotide adapters may be ligated to the fragment using T4 DNA ligase (New England Biolabs, Beverly, MA). The fragment containing the ligated adapters can be purified from the excess adapters using low melting agarose as described above. The vector pBT430 is digested, dephosphorylated with alkaline phosphatase (NEB) and deproteinized with phenol/chloroform as described above. The prepared vector pBT430 and fragment can then be ligated at 16°C for 15 hours followed by transformation into DH5 electrocompetent cells (GIBCO BRL). Transformants can be selected on agar plates containing LB media and 100 μg/mL ampicillin. Transformants containing the gene encoding the instant polypeptides are then screened for the correct orientation with respect to the T7 promoter by restriction enzyme analysis.

For high level expression, a plasmid clone with the cDNA insert in the correct orientation relative to the T7 promoter can be transformed into *E. coli* strain BL21(DE3) (Studier et al. (1986) *J. Mol. Biol. 189*:113-130). Cultures are grown in LB medium containing ampicillin (100 mg/L) at 25°C. At an optical density at 600 nm of approximately 1, IPTG (isopropylthio-β-galactoside, the inducer) can be added to a final concentration of 0.4 mM and incubation can be continued for 3 h at 25°. Cells are then harvested by centrifugation and re-suspended in 50 μL of 50 mM Tris-HCl at pH 8.0 containing 0.1 mM DTT and 0.2 mM phenyl methylsulfonyl fluoride. A small amount of 1 mm glass beads can be added and the mixture sonicated 3 times for about 5 seconds each time with a microprobe sonicator. The mixture is centrifuged and the protein concentration of the supernatant determined. One microgram of protein from the soluble

fraction of the culture can be separated by SDS-polyacrylamide gel electrophoresis. Gels can be observed for protein bands migrating at the expected molecular weight.

The above examples are provided to illustrate the invention but not to limit its scope. Other variants of the invention will be readily apparent to one of ordinary skill in the art and are encompassed by the appended claims. All publications, patents, patent applications, and computer programs cited herein are hereby incorporated by reference.

Deposits

Plasmids containing polynucleotide sequences of the invention were deposited on March 7, 2000, with the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, Virginia USA, 20110-2209, and assigned Accession Nos. PTA-1454, PTA-1455, and PTA-1456. These deposits will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. In addition, during the pendency of this patent application, access to the deposited cultures will be available to the Commissioner of Patents and Trademarks and to persons determined by the Commissioner to be entitled thereto under 37 C.F.R. §114 and 35 U.S.C. §122.

These deposits were made merely as a convenience for those of skill in the art and are not an admission that a deposit is required under 35 U.S.C. §112. All restrictions imposed by the depositor on the availability to the public of the deposited material will be irrevocably removed upon granting of a patent. However, it should be understood that the availability of a deposit does not constitute a license to practice the subject invention in derogation of patent rights granted by government action.

WHAT IS CLAIMED IS:

1. An isolated nucleic acid comprising a member selected from the group consisting of:

(a) a polynucleotide having at least 80% sequence identity, as determined by the GAP algorithm under default parameters, to a polynucleotide selected from the group consisting of SEQ ID NOS: 1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, and 49;

(b) a polynucleotide encoding a polypeptide selected from the group consisting of SEQ ID NO: 2, 6, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, and 50;

(c) a polynucleotide amplified from a *Zea mays* nucleic acid library using primers which selectively hybridize, under stringent hybridization conditions, to loci within a polynucleotide selected from the group consisting of SEQ ID NOS: 1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, and 49;

(d) a polynucleotide which selectively hybridizes, under stringent hybridization conditions and a wash in 0.1X SSC at 65°C, to a polynucleotide selected from the group consisting of SEQ ID NOS: 1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, and 49;

(e) a polynucleotide selected from the group consisting of SEQ ID NOS: 1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, and 49;

(f) a polynucleotide which is complementary to a polynucleotide of (a), (b), (c), (d), or (e); and

(g) a polynucleotide comprising at least 25 contiguous nucleotides from a polynucleotide of (a), (b), (c), (d), (e), or (f).

2. A recombinant expression cassette, comprising a nucleic acid of claim 1 operably linked, in sense or anti-sense orientation, to a promoter.

3. A host cell comprising the recombinant expression cassette of claim 2.

4. A transgenic plant comprising a recombinant expression cassette of claim 2.

5. The transgenic plant of claim 4 wherein the plant is a monocot.

6. The transgenic plant of claim 4 wherein the plant is selected from the group consisting of: maize, soybean, sunflower, sorghum, canola, wheat, alfalfa, cotton, rice, barley, and millet.

7. A transgenic seed from the transgenic plant of claim 4.

8. A method of modulating the level of cell cycle gene activity in a plant cell capable of plant regeneration, comprising:
(a) transforming the plant cell with a recombinant expression cassette comprising a cell cycle polynucleotide of claim 1 operably linked to a promoter;
(b) culturing the transformed plant cell; and
(c) inducing expression of said polynucleotide for a time sufficient to modulate the level of cell cycle gene activity in said transformed plant cell.

9. The method of claim 8 wherein a plant is regenerated from the transformed plant cell.

10. The method of claim 9 wherein the plant is selected from the group consisting of: maize, soybean, sunflower, sorghum, canola, wheat, alfalfa, cotton, rice, barley, and millet.

11. The method of claim 8 wherein the promoter is a tissue-preferred promoter.

12. The method of claim 8 wherein the level of cell cycle gene activity is increased.

13. The method of claim 8 wherein the cell cycle gene is selected from the group consisting of cyclins and cyclin-dependent kinases.

14. The method of claim 8 wherein the cell cycle polynucleotide is amplified from a *Zea mays* nucleic acid library using primers which selectively hybridize, under stringent hybridization conditions, to loci within a polynucleotide selected from the group consisting of SEQ ID NOS: 1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, and 49.

15. The method of claim 8 wherein the cell cycle polynucleotide is selected from the group consisting of SEQ ID NOS: 1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, and 49.

16. An isolated protein comprising a member selected from the group consisting of:

(a) a polypeptide of at least 20 contiguous amino acids from a polypeptide of SEQ ID NO: 2, 6, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, or 50;

(b) a polypeptide of SEQ ID NO: 2, 6, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, or 50;

(c) a polypeptide having at least 80% sequence similarity to, and having at least one epitope in common with, a polypeptide of SEQ ID NO: 2, 6, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, or 50, wherein said sequence similarity is determined using GAP under default parameters;

(d) a polypeptide encoded by a polynucleotide amplified from a *Zea mays* nucleic acid library using primers which selectively hybridize, under stringent hybridization conditions, to loci within a polynucleotide selected from the group consisting of SEQ ID NOS: 1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, and 49;

(e) a polypeptide encoded by a polynucleotide which selectively hybridizes, under stringent hybridization conditions and a wash in 0.1X SSC at 65°C, to a polynucleotide selected from the group consisting of SEQ ID NOS: 1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, and 49; and

(f) a polypeptide encoded by a polynucleotide selected from the group consisting of SEQ ID NOS: 1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, and 49.

5

SEQUENCE LISTING

<110> Pioneer Hi-Bred International, Inc.

<120> Cell Cycle Genes and Methods of Use

<130> 1081-PCT

<150> US 60/130,849
<151> 1999-04-22

<160> 52

<170> FastSEQ for Windows Version 3.0

<210> 1
<211> 1575
<212> DNA
<213> Zea mays

<220>
<221> CDS
<222> (1)...(1572)

<400> 1

```
atg tgg aaa ctc aat ccc ttt gga ggt aaa gca cag agt ggg ttg gat      48
Met Trp Lys Leu Asn Pro Phe Gly Gly Lys Ala Gln Ser Gly Leu Asp
 1               5                  10                  15

ggt cgt act att gat gtt gga aat ttg aag atc aca gta caa aat gcc      96
Gly Arg Thr Ile Asp Val Gly Asn Leu Lys Ile Thr Val Gln Asn Ala
            20                  25                  30

att gca caa gga ggt ttc tct tgt gta tat ctg gca tgt gac aca gta     144
Ile Ala Gln Gly Gly Phe Ser Cys Val Tyr Leu Ala Cys Asp Thr Val
        35                  40                  45

cat cca tca aag atg tat gca cta aag cac att att tgc aat gac tca     192
His Pro Ser Lys Met Tyr Ala Leu Lys His Ile Ile Cys Asn Asp Ser
    50                  55                  60

gaa tcg ctt gat ctt gtc atg aag gag atc cag gtt atg aac ctc ctc     240
Glu Ser Leu Asp Leu Val Met Lys Glu Ile Gln Val Met Asn Leu Leu
65                  70                  75                  80

aaa ggg cat cct aat gtg gtc aca ctt gtt gcc cat gat gtt ttt gac     288
Lys Gly His Pro Asn Val Val Thr Leu Val Ala His Asp Val Phe Asp
                85                  90                  95

atg ggt cgt aca aaa gag gca ctt ctt gta atg gag ttc tgt gag aag     336
Met Gly Arg Thr Lys Glu Ala Leu Leu Val Met Glu Phe Cys Glu Lys
            100                 105                 110

tct ttg gtt agt gca atg gag agc aga ggt agt ggg tac tat gag gag     384
Ser Leu Val Ser Ala Met Glu Ser Arg Gly Ser Gly Tyr Tyr Glu Glu
        115                 120                 125

aag aag gtg ctt ttg att ttc aga gat gtc tgc aat gct gcc ttt gct     432
Lys Lys Val Leu Leu Ile Phe Arg Asp Val Cys Asn Ala Ala Phe Ala
    130                 135                 140
```

```
atg cat gga caa tca ccg cca att gca cat agg gat ctg aaa gct gaa      480
Met His Gly Gln Ser Pro Pro Ile Ala His Arg Asp Leu Lys Ala Glu
145                 150                 155                 160

aat gtt ctt ctt ggt tgt gac ggt gta tgg aaa ata tgt gat ttt gga      528
Asn Val Leu Leu Gly Cys Asp Gly Val Trp Lys Ile Cys Asp Phe Gly
                165                 170                 175

agc aca tca act aat cac aaa tgc ttt aac aaa cca gaa gag atg ggc      576
Ser Thr Ser Thr Asn His Lys Cys Phe Asn Lys Pro Glu Glu Met Gly
            180                 185                 190

att gag gaa gat gtc atc agg aag cat aca acc cca gcc tac agg ccc      624
Ile Glu Glu Asp Val Ile Arg Lys His Thr Thr Pro Ala Tyr Arg Pro
        195                 200                 205

cca gag atg tgg gat ctc tac aga aga gaa gtt att agc gaa aaa gtt      672
Pro Glu Met Trp Asp Leu Tyr Arg Arg Glu Val Ile Ser Glu Lys Val
    210                 215                 220

gac att tgg gct ttg ggg tgc ctt cta tat aaa ata tgc tac ttc aag      720
Asp Ile Trp Ala Leu Gly Cys Leu Leu Tyr Lys Ile Cys Tyr Phe Lys
225                 230                 235                 240

tct gca ttt gat gga gaa tcc aag ttg cag ata ctc aat ggc aat tat      768
Ser Ala Phe Asp Gly Glu Ser Lys Leu Gln Ile Leu Asn Gly Asn Tyr
                245                 250                 255

cgt atc cct gag caa ccc aag tat agc act gct gtc aca ggg ttg atc      816
Arg Ile Pro Glu Gln Pro Lys Tyr Ser Thr Ala Val Thr Gly Leu Ile
            260                 265                 270

aaa gat atg ctg gaa gcc tct cca aat tcc aga cca gac atc aca cag      864
Lys Asp Met Leu Glu Ala Ser Pro Asn Ser Arg Pro Asp Ile Thr Gln
        275                 280                 285

gtc tgg ttt cgt gtt aat gag cta ctg ccg ctt gag tta caa aaa agt      912
Val Trp Phe Arg Val Asn Glu Leu Leu Pro Leu Glu Leu Gln Lys Ser
    290                 295                 300

tta cct gac gga cct tcg cca gcc gtt tcc ttg agt ttg caa gat gaa      960
Leu Pro Asp Gly Pro Ser Pro Ala Val Ser Leu Ser Leu Gln Asp Glu
305                 310                 315                 320

ggt gct cat aaa agg gca cat gtg atg cct aga agg aac cct cct cca     1008
Gly Ala His Lys Arg Ala His Val Met Pro Arg Arg Asn Pro Pro Pro
                325                 330                 335

cct cca aga gag caa tct aat agt tct cta tcg cat gga agc tca aag     1056
Pro Pro Arg Glu Gln Ser Asn Ser Ser Leu Ser His Gly Ser Ser Lys
            340                 345                 350

gca gga gat gca ccc tta ggt gca ttt tgg gca aca cag cat gca caa     1104
Ala Gly Asp Ala Pro Leu Gly Ala Phe Trp Ala Thr Gln His Ala Gln
        355                 360                 365

ggt gct caa gct gca gat aat agg aat cct ttg ttt gac gag gag cca     1152
Gly Ala Gln Ala Ala Asp Asn Arg Asn Pro Leu Phe Asp Glu Glu Pro
    370                 375                 380

agt aag gta tca ttg tca tca aag cag aac caa agt cgg gtg gac acc     1200
```

```
Ser Lys Val Ser Leu Ser Ser Lys Gln Asn Gln Ser Arg Val Asp Thr
385                 390                 395                 400

agc atc agt att cct ggt gat agg cat gat cat tct ggt cag acg tca  1248
Ser Ile Ser Ile Pro Gly Asp Arg His Asp His Ser Gly Gln Thr Ser
                405                 410                 415

cga ata agc aaa aca cca aat aac tcc ttg tcc aat gat ggt ttc aga  1296
Arg Ile Ser Lys Thr Pro Asn Asn Ser Leu Ser Asn Asp Gly Phe Arg
            420                 425                 430

ggt gtc tct gac acg gag ata cat aat tct gtg aaa act aaa gcg caa  1344
Gly Val Ser Asp Thr Glu Ile His Asn Ser Val Lys Thr Lys Ala Gln
        435                 440                 445

caa cct caa ccc aag cca aaa tgt gat aaa gat cca ttc aac atc ttt  1392
Gln Pro Gln Pro Lys Pro Lys Cys Asp Lys Asp Pro Phe Asn Ile Phe
    450                 455                 460

gtt gca gat ttt gac act cac aat ctc aac att gcc gtt ggt aag gca  1440
Val Ala Asp Phe Asp Thr His Asn Leu Asn Ile Ala Val Gly Lys Ala
465                 470                 475                 480

tcc gaa ctt gaa ctt gaa gtg tcc agt ctg aag gag cag ttg aag aaa  1488
Ser Glu Leu Glu Leu Glu Val Ser Ser Leu Lys Glu Gln Leu Lys Lys
                485                 490                 495

acc aca tta gaa aag gct gag atg aca gcc aag tat gaa agt tat ctg  1536
Thr Thr Leu Glu Lys Ala Glu Met Thr Ala Lys Tyr Glu Ser Tyr Leu
            500                 505                 510

caa tct gcc gat cac agc gtc agg aga tcc aag agc tga              1575
Gln Ser Ala Asp His Ser Val Arg Arg Ser Lys Ser
        515                 520
```

<210> 2
<211> 524
<212> PRT
<213> Zea mays

<400> 2

```
Met Trp Lys Leu Asn Pro Phe Gly Gly Lys Ala Gln Ser Gly Leu Asp
 1               5                  10                  15
Gly Arg Thr Ile Asp Val Gly Asn Leu Lys Ile Thr Val Gln Asn Ala
            20                  25                  30
Ile Ala Gln Gly Gly Phe Ser Cys Val Tyr Leu Ala Cys Asp Thr Val
        35                  40                  45
His Pro Ser Lys Met Tyr Ala Leu Lys His Ile Ile Cys Asn Asp Ser
    50                  55                  60
Glu Ser Leu Asp Leu Val Met Lys Glu Ile Gln Val Met Asn Leu Leu
65                  70                  75                  80
Lys Gly His Pro Asn Val Val Thr Leu Val Ala His Asp Val Phe Asp
                85                  90                  95
Met Gly Arg Thr Lys Glu Ala Leu Leu Val Met Glu Phe Cys Glu Lys
            100                 105                 110
Ser Leu Val Ser Ala Met Glu Ser Arg Gly Ser Gly Tyr Tyr Glu Glu
        115                 120                 125
Lys Lys Val Leu Leu Ile Phe Arg Asp Val Cys Asn Ala Ala Phe Ala
    130                 135                 140
Met His Gly Gln Ser Pro Pro Ile Ala His Arg Asp Leu Lys Ala Glu
145                 150                 155                 160
```

Asn Val Leu Leu Gly Cys Asp Gly Val Trp Lys Ile Cys Asp Phe Gly
165 170 175
Ser Thr Ser Thr Asn His Lys Cys Phe Asn Lys Pro Glu Glu Met Gly
180 185 190
Ile Glu Glu Asp Val Ile Arg Lys His Thr Thr Pro Ala Tyr Arg Pro
195 200 205
Pro Glu Met Trp Asp Leu Tyr Arg Arg Glu Val Ile Ser Glu Lys Val
210 215 220
Asp Ile Trp Ala Leu Gly Cys Leu Leu Tyr Lys Ile Cys Tyr Phe Lys
225 230 235 240
Ser Ala Phe Asp Gly Glu Ser Lys Leu Gln Ile Leu Asn Gly Asn Tyr
245 250 255
Arg Ile Pro Glu Gln Pro Lys Tyr Ser Thr Ala Val Thr Gly Leu Ile
260 265 270
Lys Asp Met Leu Glu Ala Ser Pro Asn Ser Arg Pro Asp Ile Thr Gln
275 280 285
Val Trp Phe Arg Val Asn Glu Leu Leu Pro Leu Glu Leu Gln Lys Ser
290 295 300
Leu Pro Asp Gly Pro Ser Pro Ala Val Ser Leu Ser Leu Gln Asp Glu
305 310 315 320
Gly Ala His Lys Arg Ala His Val Met Pro Arg Arg Asn Pro Pro Pro
325 330 335
Pro Pro Arg Glu Gln Ser Asn Ser Ser Leu Ser His Gly Ser Ser Lys
340 345 350
Ala Gly Asp Ala Pro Leu Gly Ala Phe Trp Ala Thr Gln His Ala Gln
355 360 365
Gly Ala Gln Ala Ala Asp Asn Arg Asn Pro Leu Phe Asp Glu Glu Pro
370 375 380
Ser Lys Val Ser Leu Ser Ser Lys Gln Asn Gln Ser Arg Val Asp Thr
385 390 395 400
Ser Ile Ser Ile Pro Gly Asp Arg His Asp His Ser Gly Gln Thr Ser
405 410 415
Arg Ile Ser Lys Thr Pro Asn Asn Ser Leu Ser Asn Asp Gly Phe Arg
420 425 430
Gly Val Ser Asp Thr Glu Ile His Asn Ser Val Lys Thr Lys Ala Gln
435 440 445
Gln Pro Gln Pro Lys Pro Lys Cys Asp Lys Asp Pro Phe Asn Ile Phe
450 455 460
Val Ala Asp Phe Asp Thr His Asn Leu Asn Ile Ala Val Gly Lys Ala
465 470 475 480
Ser Glu Leu Glu Leu Glu Val Ser Ser Leu Lys Glu Gln Leu Lys Lys
485 490 495
Thr Thr Leu Glu Lys Ala Glu Met Thr Ala Lys Tyr Glu Ser Tyr Leu
500 505 510
Gln Ser Ala Asp His Ser Val Arg Arg Ser Lys Ser
515 520

<210> 3
<211> 20
<212> DNA
<213> Zea mays

<400> 3
atgtggaaac tcaatccctt 20

<210> 4
<211> 20
<212> DNA
<213> Zea mays

<400> 4
tcagctcttg gatctcctga 20

<210> 5
<211> 1204
<212> DNA
<213> Zea mays

<220>
<221> CDS
<222> (35)...(1009)

<400> 5

```
gtcgacccac gcgtccgacc ggatctgcgg cggg atg gcg acg atc cag aac aag      55
                                      Met Ala Thr Ile Gln Asn Lys
                                       1               5

ccg act ccg acg gcg ccg tcc acg acg acg ggc ggc gga ctg cgc gcc      103
Pro Thr Pro Thr Ala Pro Ser Thr Thr Thr Gly Gly Gly Leu Arg Ala
        10                  15                  20

atg gat ctg tac gat aag ctg gag aag gtg ggg gag ggc acg tac ggg      151
Met Asp Leu Tyr Asp Lys Leu Glu Lys Val Gly Glu Gly Thr Tyr Gly
    25                  30                  35

aag gtg tac aag gcc cgg gag aag gcg acg ggc cgg atc gtg gcg ctc      199
Lys Val Tyr Lys Ala Arg Glu Lys Ala Thr Gly Arg Ile Val Ala Leu
40                  45                  50                  55

aag aag acg cgg ctc ccc gag gac gac gag ggc gtg ccg ccc acc gcg      247
Lys Lys Thr Arg Leu Pro Glu Asp Asp Glu Gly Val Pro Pro Thr Ala
                60                  65                  70

ctg cgg gag gtc tcg ctg ctg cgg atg ctg tcg cag gac ccg cac gtg      295
Leu Arg Glu Val Ser Leu Leu Arg Met Leu Ser Gln Asp Pro His Val
            75                  80                  85

gtc cgc ctg ctc gac ctc aag cag ggc gtc aac aag gag ggg cag acc      343
Val Arg Leu Leu Asp Leu Lys Gln Gly Val Asn Lys Glu Gly Gln Thr
        90                  95                  100

atc ctg tac ctc gtc ttc gag tac atg gac acc gac ctg aag aag ttc      391
Ile Leu Tyr Leu Val Phe Glu Tyr Met Asp Thr Asp Leu Lys Lys Phe
    105                 110                 115

atc cgg ggc tat cgc gcc aac cac gag aag atc ccc gca caa acc gtc      439
Ile Arg Gly Tyr Arg Ala Asn His Glu Lys Ile Pro Ala Gln Thr Val
120                 125                 130                 135

aag atc ctg atg tac cag ctg tgc aag ggc gtg ggt ttt gtt cac ggc      487
Lys Ile Leu Met Tyr Gln Leu Cys Lys Gly Val Gly Phe Val His Gly
                140                 145                 150

cgc ggg gtg ctc cac cgt gat ctg aaa ccg cac aac ctg ctc atg gac      535
Arg Gly Val Leu His Arg Asp Leu Lys Pro His Asn Leu Leu Met Asp
            155                 160                 165

cgc aag acc atg gcg ctc aag atc gct gac ctc gga ctc agc cgt gcc      583
Arg Lys Thr Met Ala Leu Lys Ile Ala Asp Leu Gly Leu Ser Arg Ala
        170                 175                 180

atc acc gtc cct gta aag aag tac acg cac gag att ctg acg ctg tgg      631
Ile Thr Val Pro Val Lys Lys Tyr Thr His Glu Ile Leu Thr Leu Trp
    185                 190                 195
```

```
tac agg gcg ccc gag gtt ctt ctt ggc gcc aca cac tac tcc acg ccg      679
Tyr Arg Ala Pro Glu Val Leu Leu Gly Ala Thr His Tyr Ser Thr Pro
200                 205                 210                 215

gtt gac att tgg tcc gtg ggc tgc ata ttt gct gag ttg gtt act aat      727
Val Asp Ile Trp Ser Val Gly Cys Ile Phe Ala Glu Leu Val Thr Asn
                220                 225                 230

cag cca ctt ttc cct ggc gac tcc gag ttg cag cag ctc ctc cac atc      775
Gln Pro Leu Phe Pro Gly Asp Ser Glu Leu Gln Gln Leu Leu His Ile
            235                 240                 245

ttc aag ttg ctg ggc acc cca aat gag gag atg tgg cca gga gta ggc      823
Phe Lys Leu Leu Gly Thr Pro Asn Glu Glu Met Trp Pro Gly Val Gly
        250                 255                 260

aag ctg ccg aac tgg cac gtg tac ccc cag tgg aag ccc act aag ctg      871
Lys Leu Pro Asn Trp His Val Tyr Pro Gln Trp Lys Pro Thr Lys Leu
    265                 270                 275

tcc act ctt gtc cct ggt ctt gac tcc gat ggc tat gat tta ctt gag      919
Ser Thr Leu Val Pro Gly Leu Asp Ser Asp Gly Tyr Asp Leu Leu Glu
280                 285                 290                 295

aaa atg ctt gca tat gag cca ggg aag cgg gtc tcg gcg aag aag gcc      967
Lys Met Leu Ala Tyr Glu Pro Gly Lys Arg Val Ser Ala Lys Lys Ala
                300                 305                 310

ctg gag cac ccg tac ttc aat gat gtt aac aag gag gtg tac             1009
Leu Glu His Pro Tyr Phe Asn Asp Val Asn Lys Glu Val Tyr
            315                 320                 325

tgaagcgacc cgagtagtag ggttcatcga gagaggatca ttgtcacctt attcttcgct   1069
tagtgtgtta atatggctac cctgaagcag tacttggggt taaaaaaaaa aaaaaaaaaa   1129
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa   1189
aaaaagggcg gccgc                                                   1204
```

<210> 6
<211> 325
<212> PRT
<213> Zea mays

<400> 6

```
Met Ala Thr Ile Gln Asn Lys Pro Thr Pro Thr Ala Pro Ser Thr Thr
 1               5                  10                  15
Thr Gly Gly Gly Leu Arg Ala Met Asp Leu Tyr Asp Lys Leu Glu Lys
            20                  25                  30
Val Gly Glu Gly Thr Tyr Gly Lys Val Tyr Lys Ala Arg Glu Lys Ala
        35                  40                  45
Thr Gly Arg Ile Val Ala Leu Lys Lys Thr Arg Leu Pro Glu Asp Asp
    50                  55                  60
Glu Gly Val Pro Pro Thr Ala Leu Arg Glu Val Ser Leu Leu Arg Met
65                  70                  75                  80
Leu Ser Gln Asp Pro His Val Val Arg Leu Leu Asp Leu Lys Gln Gly
                85                  90                  95
Val Asn Lys Glu Gly Gln Thr Ile Leu Tyr Leu Val Phe Glu Tyr Met
            100                 105                 110
Asp Thr Asp Leu Lys Lys Phe Ile Arg Gly Tyr Arg Ala Asn His Glu
        115                 120                 125
Lys Ile Pro Ala Gln Thr Val Lys Ile Leu Met Tyr Gln Leu Cys Lys
    130                 135                 140
```

```
Gly Val Gly Phe Val His Gly Arg Gly Val Leu His Arg Asp Leu Lys
145                 150                 155                 160
Pro His Asn Leu Leu Met Asp Arg Lys Thr Met Ala Leu Lys Ile Ala
                165                 170                 175
Asp Leu Gly Leu Ser Arg Ala Ile Thr Val Pro Val Lys Lys Tyr Thr
            180                 185                 190
His Glu Ile Leu Thr Leu Trp Tyr Arg Ala Pro Glu Val Leu Leu Gly
        195                 200                 205
Ala Thr His Tyr Ser Thr Pro Val Asp Ile Trp Ser Val Gly Cys Ile
    210                 215                 220
Phe Ala Glu Leu Val Thr Asn Gln Pro Leu Phe Pro Gly Asp Ser Glu
225                 230                 235                 240
Leu Gln Gln Leu Leu His Ile Phe Lys Leu Leu Gly Thr Pro Asn Glu
                245                 250                 255
Glu Met Trp Pro Gly Val Gly Lys Leu Pro Asn Trp His Val Tyr Pro
            260                 265                 270
Gln Trp Lys Pro Thr Lys Leu Ser Thr Leu Val Pro Gly Leu Asp Ser
        275                 280                 285
Asp Gly Tyr Asp Leu Leu Glu Lys Met Leu Ala Tyr Glu Pro Gly Lys
    290                 295                 300
Arg Val Ser Ala Lys Lys Ala Leu Glu His Pro Tyr Phe Asn Asp Val
305                 310                 315                 320
Asn Lys Glu Val Tyr
                325
```

<210> 7
<211> 20
<212> DNA
<213> Zea mays

<400> 7
atggcgacga tccagaacaa 20

<210> 8
<211> 20
<212> DNA
<213> Zea mays

<400> 8
tcagtacacc tccttgttaa 20

<210> 9
<211> 990
<212> DNA
<213> Zea mays

<220>
<221> CDS
<222> (1)...(987)

<400> 9

```
atg gcg acg atc cag cac cag gcg aaa ccg gcg gtg gcg gcg gcg ccg      48
Met Ala Thr Ile Gln His Gln Ala Lys Pro Ala Val Ala Ala Ala Pro
 1               5                  10                  15

agc acg act act ggc ggt ggg cag cgc gcg atg gac ctc tac gag aag      96
Ser Thr Thr Thr Gly Gly Gly Gln Arg Ala Met Asp Leu Tyr Glu Lys
            20                  25                  30

ctg gag aag gtc gga gag ggg acc tac ggg aag gtg tac agg gcg cgg     144
Leu Glu Lys Val Gly Glu Gly Thr Tyr Gly Lys Val Tyr Arg Ala Arg
        35                  40                  45
```

```
gag aag gcg acg ggg cgg atc gtg gcg ctg aag aag acg cgg ctc ccc      192
Glu Lys Ala Thr Gly Arg Ile Val Ala Leu Lys Lys Thr Arg Leu Pro
     50                  55                  60

gag gac gac gag ggc gtg ccc ccc acc gcg atg cgg gag gtc tcc ttg      240
Glu Asp Asp Glu Gly Val Pro Pro Thr Ala Met Arg Glu Val Ser Leu
 65                  70                  75                  80

ctg cgg atg ctg tcg cag gac ccg cac gtg gtg cgc ctg ctg gac ctc      288
Leu Arg Met Leu Ser Gln Asp Pro His Val Val Arg Leu Leu Asp Leu
                 85                  90                  95

aag cag ggc gtg aat aag gag ggg cag acc atc ctc tac ctg gtc ttc      336
Lys Gln Gly Val Asn Lys Glu Gly Gln Thr Ile Leu Tyr Leu Val Phe
            100                 105                 110

gag tac atg gac acc gac ctc aag aag ttc atc cgg gga cac cgc agc      384
Glu Tyr Met Asp Thr Asp Leu Lys Lys Phe Ile Arg Gly His Arg Ser
        115                 120                 125

aac aac gag aag atc ccc gcg gcc acc gtc aag atc ctg atg tac cag      432
Asn Asn Glu Lys Ile Pro Ala Ala Thr Val Lys Ile Leu Met Tyr Gln
    130                 135                 140

ctc tgc aag ggc gtg gcc ttc gtc cac ggc cgc ggg gtg ctg cac cgg      480
Leu Cys Lys Gly Val Ala Phe Val His Gly Arg Gly Val Leu His Arg
145                 150                 155                 160

gac ctc aag ccg cac aac ctc ctc atg gac cgc aag acc atg gcg ctc      528
Asp Leu Lys Pro His Asn Leu Leu Met Asp Arg Lys Thr Met Ala Leu
                165                 170                 175

aag atc gcc gac ctc ggc ctc agc cgc gcc atc acc gtc ccg gtg aag      576
Lys Ile Ala Asp Leu Gly Leu Ser Arg Ala Ile Thr Val Pro Val Lys
            180                 185                 190

aag tac acc cac gag ata ctt acc ctg tgg tac agg gcc ccc gag att      624
Lys Tyr Thr His Glu Ile Leu Thr Leu Trp Tyr Arg Ala Pro Glu Ile
        195                 200                 205

ctg ctt gga gcc acg cac tac tcc acc ccg gtt gac ata tgg tcc gtt      672
Leu Leu Gly Ala Thr His Tyr Ser Thr Pro Val Asp Ile Trp Ser Val
    210                 215                 220

ggc tgc att ttc gcc gag ctg gtc act aac cag cca ctt ttc cct ggc      720
Gly Cys Ile Phe Ala Glu Leu Val Thr Asn Gln Pro Leu Phe Pro Gly
225                 230                 235                 240

gac tcg gag ctg cag cag ctg ctc cac atc ttc aag ttg ctt ggc acc      768
Asp Ser Glu Leu Gln Gln Leu Leu His Ile Phe Lys Leu Leu Gly Thr
                245                 250                 255

cca aac gaa cag gtg tgg cca ggc gtc ggc aag ctg ccc aac tgg cac      816
Pro Asn Glu Gln Val Trp Pro Gly Val Gly Lys Leu Pro Asn Trp His
            260                 265                 270

gag tac ccc cag tgg aag ccg acg aag ctg tct gct ctt gtg ccc ggc      864
Glu Tyr Pro Gln Trp Lys Pro Thr Lys Leu Ser Ala Leu Val Pro Gly
        275                 280                 285

ctc gac gct gat ggc tac gat ctt ctt gag aaa ttg ctg gaa tac gag      912
```

```
Leu Asp Ala Asp Gly Tyr Asp Leu Leu Glu Lys Leu Leu Glu Tyr Glu
    290                 295                 300

ccg gcg aag cgg atc ccc gcg aag aag gcc ctg gag cac ccc tac ttc      960
Pro Ala Lys Arg Ile Pro Ala Lys Lys Ala Leu Glu His Pro Tyr Phe
305                 310                 315                 320

                                                                     990
aaa gat gtg agg aag gga gat gcg cac tga
Lys Asp Val Arg Lys Gly Asp Ala His
                325
```

<210> 10
<211> 329
<212> PRT
<213> Zea mays

<400> 10

```
Met Ala Thr Ile Gln His Gln Ala Lys Pro Ala Val Ala Ala Ala Pro
 1               5                  10                  15
Ser Thr Thr Thr Gly Gly Gly Gln Arg Ala Met Asp Leu Tyr Glu Lys
            20                  25                  30
Leu Glu Lys Val Gly Glu Gly Thr Tyr Gly Lys Val Tyr Arg Ala Arg
        35                  40                  45
Glu Lys Ala Thr Gly Arg Ile Val Ala Leu Lys Lys Thr Arg Leu Pro
    50                  55                  60
Glu Asp Asp Glu Gly Val Pro Pro Thr Ala Met Arg Glu Val Ser Leu
65                  70                  75                  80
Leu Arg Met Leu Ser Gln Asp Pro His Val Val Arg Leu Leu Asp Leu
                85                  90                  95
Lys Gln Gly Val Asn Lys Glu Gly Gln Thr Ile Leu Tyr Leu Val Phe
            100                 105                 110
Glu Tyr Met Asp Thr Asp Leu Lys Lys Phe Ile Arg Gly His Arg Ser
        115                 120                 125
Asn Asn Glu Lys Ile Pro Ala Ala Thr Val Lys Ile Leu Met Tyr Gln
    130                 135                 140
Leu Cys Lys Gly Val Ala Phe Val His Gly Arg Gly Val Leu His Arg
145                 150                 155                 160
Asp Leu Lys Pro His Asn Leu Leu Met Asp Arg Lys Thr Met Ala Leu
                165                 170                 175
Lys Ile Ala Asp Leu Gly Leu Ser Arg Ala Ile Thr Val Pro Val Lys
            180                 185                 190
Lys Tyr Thr His Glu Ile Leu Thr Leu Trp Tyr Arg Ala Pro Glu Ile
        195                 200                 205
Leu Leu Gly Ala Thr His Tyr Ser Thr Pro Val Asp Ile Trp Ser Val
    210                 215                 220
Gly Cys Ile Phe Ala Glu Leu Val Thr Asn Gln Pro Leu Phe Pro Gly
225                 230                 235                 240
Asp Ser Glu Leu Gln Gln Leu Leu His Ile Phe Lys Leu Leu Gly Thr
                245                 250                 255
Pro Asn Glu Gln Val Trp Pro Gly Val Gly Lys Leu Pro Asn Trp His
            260                 265                 270
Glu Tyr Pro Gln Trp Lys Pro Thr Lys Leu Ser Ala Leu Val Pro Gly
        275                 280                 285
Leu Asp Ala Asp Gly Tyr Asp Leu Leu Glu Lys Leu Leu Glu Tyr Glu
    290                 295                 300
Pro Ala Lys Arg Ile Pro Ala Lys Lys Ala Leu Glu His Pro Tyr Phe
305                 310                 315                 320
Lys Asp Val Arg Lys Gly Asp Ala His
                325
```

<210> 11

<211> 20
<212> DNA
<213> Zea mays

<400> 11
atggcgacga tccagcacca 20

<210> 12
<211> 20
<212> DNA
<213> Zea mays

<400> 12
tcagtgcgca tctcccttcc 20

<210> 13
<211> 879
<212> DNA
<213> Zea mays

<220>
<221> CDS
<222> (1)...(876)

<400> 13

```
atg gac cag tac gag aag gtg gag aag atc ggg gag ggc acg tac ggg      48
Met Asp Gln Tyr Glu Lys Val Glu Lys Ile Gly Glu Gly Thr Tyr Gly
 1               5                  10                  15

gtg gtg tac aag ggc aag gac cgc cac acc aac gag acg atc gcg ctc      96
Val Val Tyr Lys Gly Lys Asp Arg His Thr Asn Glu Thr Ile Ala Leu
            20                  25                  30

aag aag atc cgc ctc gag cag gag gac gag ggc gtc ccc tcc acc gcc     144
Lys Lys Ile Arg Leu Glu Gln Glu Asp Glu Gly Val Pro Ser Thr Ala
        35                  40                  45

atc cgc gag atc tcc ctc ctc aag gag atg cag cac cgc aac atc gtc     192
Ile Arg Glu Ile Ser Leu Leu Lys Glu Met Gln His Arg Asn Ile Val
    50                  55                  60

agg ctg cag gaa gtc gtg cac aac gac aag tgc atc tac ctc gtc ttc     240
Arg Leu Gln Glu Val Val His Asn Asp Lys Cys Ile Tyr Leu Val Phe
65                  70                  75                  80

gag tac ctc gac ctc gac ctc aag aag cac atg gac tcc tcc acg gac     288
Glu Tyr Leu Asp Leu Asp Leu Lys Lys His Met Asp Ser Ser Thr Asp
                85                  90                  95

ttc aag aac cac cgc ata gtc aaa tcc ttc ctc tac cag att ctg cgg     336
Phe Lys Asn His Arg Ile Val Lys Ser Phe Leu Tyr Gln Ile Leu Arg
            100                 105                 110

ggc atc gcc tac tgc cac tcg cac cgc gtg ctc cac cgc gac ctg aag     384
Gly Ile Ala Tyr Cys His Ser His Arg Val Leu His Arg Asp Leu Lys
        115                 120                 125

ccg cag aac ctg ctg att gac cgc cgc aac aac ctc ttg aag ctc gcg     432
Pro Gln Asn Leu Leu Ile Asp Arg Arg Asn Asn Leu Leu Lys Leu Ala
    130                 135                 140

gac ttt gga ctg gcg agg gcg ttc ggc atc cct gtc cgg acg ttc act     480
```

```
Asp Phe Gly Leu Ala Arg Ala Phe Gly Ile Pro Val Arg Thr Phe Thr
145                 150                 155                 160

cat gag gtg gtg acg ctt tgg tat aga gcg cct gaa atc ctt ctc ggt      528
His Glu Val Val Thr Leu Trp Tyr Arg Ala Pro Glu Ile Leu Leu Gly
                165                 170                 175

gca agg cat tat tcc acc cct gtt gat gtg tgg tca gtt ggt tgc att      576
Ala Arg His Tyr Ser Thr Pro Val Asp Val Trp Ser Val Gly Cys Ile
            180                 185                 190

ttc gct gaa atg gtg aac cag aag gcg ctt ttt cct ggc gac tct gag      624
Phe Ala Glu Met Val Asn Gln Lys Ala Leu Phe Pro Gly Asp Ser Glu
        195                 200                 205

atc gat gag ctg ttt aag att ttc aga att ttg ggc act cca act aaa      672
Ile Asp Glu Leu Phe Lys Ile Phe Arg Ile Leu Gly Thr Pro Thr Lys
    210                 215                 220

gaa aca tgg cca ggc gtt gct tcg ttg cct gat tac aag tca act ttc      720
Glu Thr Trp Pro Gly Val Ala Ser Leu Pro Asp Tyr Lys Ser Thr Phe
225                 230                 235                 240

cca aag tgg cca cct gtg gat ctt gca acg gtg gtc ccg aca ctc gaa      768
Pro Lys Trp Pro Pro Val Asp Leu Ala Thr Val Val Pro Thr Leu Glu
                245                 250                 255

ccg tcg gga atc gat ctc cta tct aag atg ctg cgt cta gat ccc agc      816
Pro Ser Gly Ile Asp Leu Leu Ser Lys Met Leu Arg Leu Asp Pro Ser
            260                 265                 270

aag agg atc acc gcc cgc gcc gcc ctc gag cac gac tac ttc agg gac      864
Lys Arg Ile Thr Ala Arg Ala Ala Leu Glu His Asp Tyr Phe Arg Asp
        275                 280                 285

ctc gag cac gcc tag                                                  879
Leu Glu His Ala
    290
```

<210> 14
<211> 292
<212> PRT
<213> Zea mays

<400> 14

```
Met Asp Gln Tyr Glu Lys Val Glu Lys Ile Gly Glu Gly Thr Tyr Gly
 1               5                  10                  15
Val Val Tyr Lys Gly Lys Asp Arg His Thr Asn Glu Thr Ile Ala Leu
            20                  25                  30
Lys Lys Ile Arg Leu Glu Gln Glu Asp Glu Gly Val Pro Ser Thr Ala
        35                  40                  45
Ile Arg Glu Ile Ser Leu Leu Lys Glu Met Gln His Arg Asn Ile Val
    50                  55                  60
Arg Leu Gln Glu Val Val His Asn Asp Lys Cys Ile Tyr Leu Val Phe
65                  70                  75                  80
Glu Tyr Leu Asp Leu Asp Leu Lys Lys His Met Asp Ser Ser Thr Asp
                85                  90                  95
Phe Lys Asn His Arg Ile Val Lys Ser Phe Leu Tyr Gln Ile Leu Arg
            100                 105                 110
Gly Ile Ala Tyr Cys His Ser His Arg Val Leu His Arg Asp Leu Lys
        115                 120                 125
```

```
Pro Gln Asn Leu Leu Ile Asp Arg Arg Asn Asn Leu Leu Lys Leu Ala
    130                 135                 140
Asp Phe Gly Leu Ala Arg Ala Phe Gly Ile Pro Val Arg Thr Phe Thr
145                 150                 155                 160
His Glu Val Val Thr Leu Trp Tyr Arg Ala Pro Glu Ile Leu Leu Gly
                165                 170                 175
Ala Arg His Tyr Ser Thr Pro Val Asp Val Trp Ser Val Gly Cys Ile
            180                 185                 190
Phe Ala Glu Met Val Asn Gln Lys Ala Leu Phe Pro Gly Asp Ser Glu
        195                 200                 205
Ile Asp Glu Leu Phe Lys Ile Phe Arg Ile Leu Gly Thr Pro Thr Lys
    210                 215                 220
Glu Thr Trp Pro Gly Val Ala Ser Leu Pro Asp Tyr Lys Ser Thr Phe
225                 230                 235                 240
Pro Lys Trp Pro Pro Val Asp Leu Ala Thr Val Val Pro Thr Leu Glu
                245                 250                 255
Pro Ser Gly Ile Asp Leu Leu Ser Lys Met Leu Arg Leu Asp Pro Ser
            260                 265                 270
Lys Arg Ile Thr Ala Arg Ala Ala Leu Glu His Asp Tyr Phe Arg Asp
        275                 280                 285
Leu Glu His Ala
    290
```

<210> 15
<211> 20
<212> DNA
<213> Zea mays

<400> 15
atggaccagt acgagaaggt 20

<210> 16
<211> 20
<212> DNA
<213> Zea mays

<400> 16
ctaggcgtgc tcgaggtccc 20

<210> 17
<211> 993
<212> DNA
<213> Zea mays

<220>
<221> CDS
<222> (1)...(990)

<400> 17

```
atg ccc ttg ccc atc gac ggt aga cga gag cga gcg gtc tgt gtt cgt      48
Met Pro Leu Pro Ile Asp Gly Arg Arg Glu Arg Ala Val Cys Val Arg
 1               5                  10                  15

gga gca ggg gcc ggc gcc atg gag atc aac atc gtg gac aag tac gag      96
Gly Ala Gly Ala Gly Ala Met Glu Ile Asn Ile Val Asp Lys Tyr Glu
            20                  25                  30

aag ctg gag aag gtc ggg gag ggc acc tac ggc aag gtg tac aag gcg     144
Lys Leu Glu Lys Val Gly Glu Gly Thr Tyr Gly Lys Val Tyr Lys Ala
        35                  40                  45

cag gac aag gcg acg ggg cag ctg gtg gcg ctg aag aag acc cgc ctc     192
```

```
Gln Asp Lys Ala Thr Gly Gln Leu Val Ala Leu Lys Lys Thr Arg Leu
     50                  55                  60

gag atg gac gag gag ggc atc ccg ccc acg gcg ctc cgc gag atc tcc      240
Glu Met Asp Glu Glu Gly Ile Pro Pro Thr Ala Leu Arg Glu Ile Ser
 65                  70                  75                  80

ctc ctc aac ctc ctc tcc cac tcc atc tac atc gtc cgc ctc ctc gcc      288
Leu Leu Asn Leu Leu Ser His Ser Ile Tyr Ile Val Arg Leu Leu Ala
                 85                  90                  95

gtc gag cag gcc gct aag aat ggc aag ccc gtc ctc tac ctc gtc ttc      336
Val Glu Gln Ala Ala Lys Asn Gly Lys Pro Val Leu Tyr Leu Val Phe
            100                 105                 110

gag ttc ctc gac acc gat ctc aag aag tac ctc gac gtc tac cgc agg      384
Glu Phe Leu Asp Thr Asp Leu Lys Lys Tyr Leu Asp Val Tyr Arg Arg
        115                 120                 125

ggg ccc agc gcc agg ccg ctc ccc gcg aca ctc atc aag aat ttc ctg      432
Gly Pro Ser Ala Arg Pro Leu Pro Ala Thr Leu Ile Lys Asn Phe Leu
    130                 135                 140

tat cag tta tgc aaa gga gtt gca cac tgc cat agc cat ggt gtc ctt      480
Tyr Gln Leu Cys Lys Gly Val Ala His Cys His Ser His Gly Val Leu
145                 150                 155                 160

cac cgg gat tta aag cca caa aac ctg ctg gtg gac aag gag aag ggt      528
His Arg Asp Leu Lys Pro Gln Asn Leu Leu Val Asp Lys Glu Lys Gly
                165                 170                 175

ata ctg aag att gct gat ctg gga ctc ggt agg gct ttc act gtg ccc      576
Ile Leu Lys Ile Ala Asp Leu Gly Leu Gly Arg Ala Phe Thr Val Pro
            180                 185                 190

atg aaa agc tac acc cat gag att gtg acc ctt tgg tac aga gct cct      624
Met Lys Ser Tyr Thr His Glu Ile Val Thr Leu Trp Tyr Arg Ala Pro
        195                 200                 205

gaa gtt ttg ctt gga gca aca cat tac tca act ggc gtt gat atg tgg      672
Glu Val Leu Leu Gly Ala Thr His Tyr Ser Thr Gly Val Asp Met Trp
    210                 215                 220

tct gtt gga tgc atc ttt gct gaa atg gcc cga cgg cag gct ctc ttt      720
Ser Val Gly Cys Ile Phe Ala Glu Met Ala Arg Arg Gln Ala Leu Phe
225                 230                 235                 240

cct ggt gac tcg gag ttg caa cag ttg ctt cac atc ttc agg ttg ctg      768
Pro Gly Asp Ser Glu Leu Gln Gln Leu Leu His Ile Phe Arg Leu Leu
                245                 250                 255

gga aca cct act gag gaa cag tgg cct gga gtg agt gat tta agg gac      816
Gly Thr Pro Thr Glu Glu Gln Trp Pro Gly Val Ser Asp Leu Arg Asp
            260                 265                 270

tgg cat gag ttt cca cag tgg aag ccc cag ggt ttg gca cgt gtc gtc      864
Trp His Glu Phe Pro Gln Trp Lys Pro Gln Gly Leu Ala Arg Val Val
        275                 280                 285

cca aca ctg gaa cct gaa gga gtt gac ctt cta tcg aaa atg ctt cag      912
Pro Thr Leu Glu Pro Glu Gly Val Asp Leu Leu Ser Lys Met Leu Gln
    290                 295                 300
```

```
ctt gat cct tca aac agg ata tca gct tta gct gcg atg gag cat ccc      960
Leu Asp Pro Ser Asn Arg Ile Ser Ala Leu Ala Ala Met Glu His Pro
305                 310                 315                 320

tac ttt aac agc ctt gac aag tcc cag ttc tag                          993
Tyr Phe Asn Ser Leu Asp Lys Ser Gln Phe
                325                 330
```

<210> 18
<211> 330
<212> PRT
<213> Zea mays

<400> 18

```
Met Pro Leu Pro Ile Asp Gly Arg Arg Glu Arg Ala Val Cys Val Arg
 1               5                  10                  15
Gly Ala Gly Ala Gly Ala Met Glu Ile Asn Ile Val Asp Lys Tyr Glu
            20                  25                  30
Lys Leu Glu Lys Val Gly Glu Gly Thr Tyr Gly Lys Val Tyr Lys Ala
        35                  40                  45
Gln Asp Lys Ala Thr Gly Gln Leu Val Ala Leu Lys Lys Thr Arg Leu
    50                  55                  60
Glu Met Asp Glu Glu Gly Ile Pro Pro Thr Ala Leu Arg Glu Ile Ser
65                  70                  75                  80
Leu Leu Asn Leu Leu Ser His Ser Ile Tyr Ile Val Arg Leu Leu Ala
                85                  90                  95
Val Glu Gln Ala Ala Lys Asn Gly Lys Pro Val Leu Tyr Leu Val Phe
            100                 105                 110
Glu Phe Leu Asp Thr Asp Leu Lys Lys Tyr Leu Asp Val Tyr Arg Arg
        115                 120                 125
Gly Pro Ser Ala Arg Pro Leu Pro Ala Thr Leu Ile Lys Asn Phe Leu
    130                 135                 140
Tyr Gln Leu Cys Lys Gly Val Ala His Cys His Ser His Gly Val Leu
145                 150                 155                 160
His Arg Asp Leu Lys Pro Gln Asn Leu Leu Val Asp Lys Glu Lys Gly
                165                 170                 175
Ile Leu Lys Ile Ala Asp Leu Gly Leu Gly Arg Ala Phe Thr Val Pro
            180                 185                 190
Met Lys Ser Tyr Thr His Glu Ile Val Thr Leu Trp Tyr Arg Ala Pro
        195                 200                 205
Glu Val Leu Leu Gly Ala Thr His Tyr Ser Thr Gly Val Asp Met Trp
    210                 215                 220
Ser Val Gly Cys Ile Phe Ala Glu Met Ala Arg Arg Gln Ala Leu Phe
225                 230                 235                 240
Pro Gly Asp Ser Glu Leu Gln Gln Leu Leu His Ile Phe Arg Leu Leu
                245                 250                 255
Gly Thr Pro Thr Glu Glu Gln Trp Pro Gly Val Ser Asp Leu Arg Asp
            260                 265                 270
Trp His Glu Phe Pro Gln Trp Lys Pro Gln Gly Leu Ala Arg Val Val
        275                 280                 285
Pro Thr Leu Glu Pro Glu Gly Val Asp Leu Leu Ser Lys Met Leu Gln
    290                 295                 300
Leu Asp Pro Ser Asn Arg Ile Ser Ala Leu Ala Ala Met Glu His Pro
305                 310                 315                 320
Tyr Phe Asn Ser Leu Asp Lys Ser Gln Phe
                325                 330
```

<210> 19
<211> 20
<212> DNA

<213> Zea mays

<400> 19
atgcccttgc ccatcgacgg 20

<210> 20
<211> 20
<212> DNA
<213> Zea mays

<400> 20
ctagaactgg gacttgtcaa 20

<210> 21
<211> 1980
<212> DNA
<213> Zea mays

<220>
<221> CDS
<222> (273)...(1619)

<400> 21
gtcgacccac gcgtccgctc ctcgtcccct cgagaactgt accatcacgc gccgcctttg 60
attgagcagg gactgggaga gaggagaaga gagggtgaaa gggggcaaga aaacgcttct 120
tgttgccttt gtgcttcagc ggatccgcgg tggctgcgca cgccggcggg gaagaggagt 180
cgcgggcggc gcgcctccag tctttttttcg cctccgtgct ctccgggctc ttcggccagg 240
ccgaggccga ggacgaggac gaggacgaag ag atg ccc acg cgc aac cac aac 293
Met Pro Thr Arg Asn His Asn
1 5

gct gcc gcg gct gcc gct ccg cag cat cac cat cac aac cga ggt ggt 341
Ala Ala Ala Ala Ala Ala Pro Gln His His His His Asn Arg Gly Gly
10 15 20

gct ccc gct ctt gga aag agc aag gcc gtg ccc ggc cga gct gat gcc 389
Ala Pro Ala Leu Gly Lys Ser Lys Ala Val Pro Gly Arg Ala Asp Ala
25 30 35

atg aac cgg cga gcc ccc ctc ggc gat att ggc aac ctc gtc agc gtc 437
Met Asn Arg Arg Ala Pro Leu Gly Asp Ile Gly Asn Leu Val Ser Val
40 45 50 55

cgc cca gcc gaa ggg aaa cct cag ctg cag gag cag atc aat cgc ccc 485
Arg Pro Ala Glu Gly Lys Pro Gln Leu Gln Glu Gln Ile Asn Arg Pro
60 65 70

atc acg cga agc ttc ggc gct cag ctc gtg aag aac gtg cag gcg aat 533
Ile Thr Arg Ser Phe Gly Ala Gln Leu Val Lys Asn Val Gln Ala Asn
75 80 85

gcc gca atc aag aat gcc gca atc cta ccc gcg agg cat gcg ccg agg 581
Ala Ala Ile Lys Asn Ala Ala Ile Leu Pro Ala Arg His Ala Pro Arg
90 95 100

cag gaa agg aag gct cct gcc aag cag ccg cct cct gag gat gtc ata 629
Gln Glu Arg Lys Ala Pro Ala Lys Gln Pro Pro Pro Glu Asp Val Ile
105 110 115

gta ctc agc tcc gac tct gaa cag agc agg acg cag ttg gag agc agc 677
Val Leu Ser Ser Asp Ser Glu Gln Ser Arg Thr Gln Leu Glu Ser Ser
120 125 130 135

```
gct agc tcc gtc cgg tcg agg aag aag gtc atc aac acc ctt tct tct      725
Ala Ser Ser Val Arg Ser Arg Lys Lys Val Ile Asn Thr Leu Ser Ser
                140                 145                 150

gtg ctc tcg gct cgc tca aag gct gcc tgt gga atc act gat aag aga      773
Val Leu Ser Ala Arg Ser Lys Ala Ala Cys Gly Ile Thr Asp Lys Arg
            155                 160                 165

cgg caa gta gca gtg atc gaa gac atc gac aag ttg gac gtc aac aat      821
Arg Gln Val Ala Val Ile Glu Asp Ile Asp Lys Leu Asp Val Asn Asn
        170                 175                 180

gag ctc gca gtt gtg gaa tac att gag gac atc tac acg ttc tac aag      869
Glu Leu Ala Val Val Glu Tyr Ile Glu Asp Ile Tyr Thr Phe Tyr Lys
    185                 190                 195

att gct cag cat gac aga cgg cca tgt gat tat ata gac acc caa gtc      917
Ile Ala Gln His Asp Arg Arg Pro Cys Asp Tyr Ile Asp Thr Gln Val
200                 205                 210                 215

gag atc aac cct aag atg agg gct atc ctg gct ggt tgg ata att gaa      965
Glu Ile Asn Pro Lys Met Arg Ala Ile Leu Ala Gly Trp Ile Ile Glu
                220                 225                 230

gta cac cac aag ttc gag ctg atg ccg gaa act ctc tac ttg acc atg     1013
Val His His Lys Phe Glu Leu Met Pro Glu Thr Leu Tyr Leu Thr Met
            235                 240                 245

tac atc atc gat cag tac ctc tcg ctg caa cca gtc ctg cga agg gag     1061
Tyr Ile Ile Asp Gln Tyr Leu Ser Leu Gln Pro Val Leu Arg Arg Glu
        250                 255                 260

ctg cag ctg gtc ggt gtt tca gct atg ctg atc gcc tgc aag tac gag     1109
Leu Gln Leu Val Gly Val Ser Ala Met Leu Ile Ala Cys Lys Tyr Glu
    265                 270                 275

gag att tgg gcc cca gag gtg aac gat ttc att ctt ata tca gac agt     1157
Glu Ile Trp Ala Pro Glu Val Asn Asp Phe Ile Leu Ile Ser Asp Ser
280                 285                 290                 295

gca tac agc agg gag cag atc ctt tcg atg gag aag gga atc ctg aat     1205
Ala Tyr Ser Arg Glu Gln Ile Leu Ser Met Glu Lys Gly Ile Leu Asn
                300                 305                 310

agc ctg gag tgg aac ctc act gtc cct aca gta tac atg ttc ctt gtt     1253
Ser Leu Glu Trp Asn Leu Thr Val Pro Thr Val Tyr Met Phe Leu Val
            315                 320                 325

cgt ttt ctg aag gcg gca gcc ttg ggc aac aaa gtt gag aaa gag atg     1301
Arg Phe Leu Lys Ala Ala Ala Leu Gly Asn Lys Val Glu Lys Glu Met
        330                 335                 340

gag aat atg gtc ttc ttc ttc gct gaa ctg gcg ctg atg cag tac ggc     1349
Glu Asn Met Val Phe Phe Phe Ala Glu Leu Ala Leu Met Gln Tyr Gly
    345                 350                 355

ttg gtg acg cgg ctg cct tcg ctg gtc gct gct tcg gtt gtc tac gca     1397
Leu Val Thr Arg Leu Pro Ser Leu Val Ala Ala Ser Val Val Tyr Ala
360                 365                 370                 375

gcc agg ctc act ctc aag agg gct ccc ctc tgg acc gac acc ctc aag     1445
```

```
Ala Arg Leu Thr Leu Lys Arg Ala Pro Leu Trp Thr Asp Thr Leu Lys
                380                 385                 390

cac cac acg ggc ttc aga gag tca gag aca gag cta atc gag tgc acg     1493
His His Thr Gly Phe Arg Glu Ser Glu Thr Glu Leu Ile Glu Cys Thr
            395                 400                 405

aag ttg ctg gtc agc gca cac tcg tcc gcc gct gac agc aag ctg agg     1541
Lys Leu Leu Val Ser Ala His Ser Ser Ala Ala Asp Ser Lys Leu Arg
        410                 415                 420

tct gta tac aag aag tat tcc agt gag cag ttc gga ggt gtc gcg ctt     1589
Ser Val Tyr Lys Lys Tyr Ser Ser Glu Gln Phe Gly Gly Val Ala Leu
    425                 430                 435

cgc cca ccc gca gcc gca gtg gag atc aag taaacgaaga taacttcgtg       1639
Arg Pro Pro Ala Ala Ala Val Glu Ile Lys
440                 445

ctagtgtttg ttagtataat gtaatggaag gcgagcctga cattcgctag ggtctatccg   1699
tctatgtttg cttgtcgggg tagtcaagaa agtttgaatg cttgaggcag ccgtttgttt   1759
ggtaggctgt tctcagagtc caggtgttat ctgacttgtg atggcatgtt gtgtttcatg   1819
tgatggatat cattagaatc ctaatctagt acgatatcta ttcatgaata taaatggtat   1879
ttattcattt gttaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa   1939
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa agggcggccg c                       1980
```

<210> 22
<211> 449
<212> PRT
<213> Zea mays

<400> 22

```
Met Pro Thr Arg Asn His Asn Ala Ala Ala Ala Ala Pro Gln His
 1               5                  10                  15
His His His Asn Arg Gly Gly Ala Pro Ala Leu Gly Lys Ser Lys Ala
            20                  25                  30
Val Pro Gly Arg Ala Asp Ala Met Asn Arg Arg Ala Pro Leu Gly Asp
        35                  40                  45
Ile Gly Asn Leu Val Ser Val Arg Pro Ala Glu Gly Lys Pro Gln Leu
    50                  55                  60
Gln Glu Gln Ile Asn Arg Pro Ile Thr Arg Ser Phe Gly Ala Gln Leu
65                  70                  75                  80
Val Lys Asn Val Gln Ala Asn Ala Ala Ile Lys Asn Ala Ala Ile Leu
                85                  90                  95
Pro Ala Arg His Ala Pro Arg Gln Glu Arg Lys Ala Pro Ala Lys Gln
            100                 105                 110
Pro Pro Pro Glu Asp Val Ile Val Leu Ser Ser Asp Ser Glu Gln Ser
        115                 120                 125
Arg Thr Gln Leu Glu Ser Ser Ala Ser Ser Val Arg Ser Arg Lys Lys
    130                 135                 140
Val Ile Asn Thr Leu Ser Ser Val Leu Ser Ala Arg Ser Lys Ala Ala
145                 150                 155                 160
Cys Gly Ile Thr Asp Lys Arg Arg Gln Val Ala Val Ile Glu Asp Ile
                165                 170                 175
Asp Lys Leu Asp Val Asn Asn Glu Leu Ala Val Val Glu Tyr Ile Glu
            180                 185                 190
Asp Ile Tyr Thr Phe Tyr Lys Ile Ala Gln His Asp Arg Arg Pro Cys
        195                 200                 205
Asp Tyr Ile Asp Thr Gln Val Glu Ile Asn Pro Lys Met Arg Ala Ile
    210                 215                 220
Leu Ala Gly Trp Ile Ile Glu Val His His Lys Phe Glu Leu Met Pro
225                 230                 235                 240
```

```
Glu Thr Leu Tyr Leu Thr Met Tyr Ile Ile Asp Gln Tyr Leu Ser Leu
                245                 250                 255
Gln Pro Val Leu Arg Arg Glu Leu Gln Leu Val Gly Val Ser Ala Met
            260                 265                 270
Leu Ile Ala Cys Lys Tyr Glu Glu Ile Trp Ala Pro Glu Val Asn Asp
        275                 280                 285
Phe Ile Leu Ile Ser Asp Ser Ala Tyr Ser Arg Glu Gln Ile Leu Ser
    290                 295                 300
Met Glu Lys Gly Ile Leu Asn Ser Leu Glu Trp Asn Leu Thr Val Pro
305                 310                 315                 320
Thr Val Tyr Met Phe Leu Val Arg Phe Leu Lys Ala Ala Ala Leu Gly
                325                 330                 335
Asn Lys Val Glu Lys Glu Met Glu Asn Met Val Phe Phe Phe Ala Glu
            340                 345                 350
Leu Ala Leu Met Gln Tyr Gly Leu Val Thr Arg Leu Pro Ser Leu Val
        355                 360                 365
Ala Ala Ser Val Val Tyr Ala Ala Arg Leu Thr Leu Lys Arg Ala Pro
    370                 375                 380
Leu Trp Thr Asp Thr Leu Lys His His Thr Gly Phe Arg Glu Ser Glu
385                 390                 395                 400
Thr Glu Leu Ile Glu Cys Thr Lys Leu Leu Val Ser Ala His Ser Ser
                405                 410                 415
Ala Ala Asp Ser Lys Leu Arg Ser Val Tyr Lys Lys Tyr Ser Ser Glu
            420                 425                 430
Gln Phe Gly Gly Val Ala Leu Arg Pro Pro Ala Ala Ala Val Glu Ile
        435                 440                 445
Lys
```

<210> 23
<211> 20
<212> DNA
<213> Zea mays

<400> 23
atgcccacgc gcaaccacaa 20

<210> 24
<211> 20
<212> DNA
<213> Zea mays

<400> 24
ttacttgatc tccactgcgg 20

<210> 25
<211> 1530
<212> DNA
<213> Zea mays

<220>
<221> CDS
<222> (1)...(1527)

<400> 25

```
atg tcg acc atc gct gcc tcc cgc cgc tcg tcc tcc tcg tcc tct tcc      48
Met Ser Thr Ile Ala Ala Ser Arg Arg Ser Ser Ser Ser Ser Ser Ser
 1               5                  10                  15

gcg acc gcc aag cgc ccc gcg att gcg gac ggc ccc ggg gga ccc aag      96
Ala Thr Ala Lys Arg Pro Ala Ile Ala Asp Gly Pro Gly Gly Pro Lys
            20                  25                  30
```

```
gcg acc gcc gca cag gcc aag aag cgc gtg gcg ctc ggc aac atc acc      144
Ala Thr Ala Ala Gln Ala Lys Lys Arg Val Ala Leu Gly Asn Ile Thr
         35                  40                  45

aac gtt gcc gcg cgg gga ggg agg gct tcc gtt ggc ggc agc ctt ggg      192
Asn Val Ala Ala Arg Gly Gly Arg Ala Ser Val Gly Gly Ser Leu Gly
     50                  55                  60

aac gtc atg gcg ccc acc agc agc gcg aag ttg aat cca acg gta ccc      240
Asn Val Met Ala Pro Thr Ser Ser Ala Lys Leu Asn Pro Thr Val Pro
 65                  70                  75                  80

ttg aag aag ccg tct ttg gca aca agt gct cgg agt gtg agt tct tcc      288
Leu Lys Lys Pro Ser Leu Ala Thr Ser Ala Arg Ser Val Ser Ser Ser
                 85                  90                  95

atc cgg ggt tct gct gac aaa ccg gct tcc atc aag cca gct cca cca      336
Ile Arg Gly Ser Ala Asp Lys Pro Ala Ser Ile Lys Pro Ala Pro Pro
            100                 105                 110

gta gca cgc cat ggc agc gca aca caa agg cat aac aat gtt cct cct      384
Val Ala Arg His Gly Ser Ala Thr Gln Arg His Asn Asn Val Pro Pro
        115                 120                 125

cct aaa gtg cct acc att gcc gat gtg cca agt cgt gct cct gcc ttg      432
Pro Lys Val Pro Thr Ile Ala Asp Val Pro Ser Arg Ala Pro Ala Leu
    130                 135                 140

gtg tcc tgc acc ggc ttg gtg tct cct gga cgc tca gga gat tct gtt      480
Val Ser Cys Thr Gly Leu Val Ser Pro Gly Arg Ser Gly Asp Ser Val
145                 150                 155                 160

tca tcc gat gag acg atg tcg act tgc gac tcc atg aaa agc cca gac      528
Ser Ser Asp Glu Thr Met Ser Thr Cys Asp Ser Met Lys Ser Pro Asp
                165                 170                 175

ttt gag tac gta gac aat cag gac acc tcg atg ctg gct tcg ttg cag      576
Phe Glu Tyr Val Asp Asn Gln Asp Thr Ser Met Leu Ala Ser Leu Gln
            180                 185                 190

cga cgg aca agt gaa cat ctg cga ata tca gag gat aga gat gtt gaa      624
Arg Arg Thr Ser Glu His Leu Arg Ile Ser Glu Asp Arg Asp Val Glu
        195                 200                 205

gaa aat aag cgg aag aaa aat gcc gtg gcc cca atg gaa att gac cgc      672
Glu Asn Lys Arg Lys Lys Asn Ala Val Ala Pro Met Glu Ile Asp Arg
    210                 215                 220

atc tgt gat gtt gac agt gaa tat gag gat cca cag ctg tgc gct act      720
Ile Cys Asp Val Asp Ser Glu Tyr Glu Asp Pro Gln Leu Cys Ala Thr
225                 230                 235                 240

ctg gct tct gac att tac atg cac ttg cga gag gct gag acg aag aaa      768
Leu Ala Ser Asp Ile Tyr Met His Leu Arg Glu Ala Glu Thr Lys Lys
                245                 250                 255

aga ccg tca act gat ttc atg gaa acg att caa aaa gat gtc aac cca      816
Arg Pro Ser Thr Asp Phe Met Glu Thr Ile Gln Lys Asp Val Asn Pro
            260                 265                 270

agc atg aga gca att ttg ata gac tgg ctc gtg gaa gtt gct gaa gaa      864
```

```
Ser Met Arg Ala Ile Leu Ile Asp Trp Leu Val Glu Val Ala Glu Glu
        275                 280                 285

tat cgt ctt gtt cct gat act ttg tac ctg aca gtc aac tac att gac      912
Tyr Arg Leu Val Pro Asp Thr Leu Tyr Leu Thr Val Asn Tyr Ile Asp
    290                 295                 300

cgt tat ctt tct ggc aac gaa atc agc cgc caa cgg ctg caa tta ctt      960
Arg Tyr Leu Ser Gly Asn Glu Ile Ser Arg Gln Arg Leu Gln Leu Leu
305                 310                 315                 320

ggt gtt gca tgc atg ctt ata gct gca aaa tac gag gag ata tgt gca     1008
Gly Val Ala Cys Met Leu Ile Ala Ala Lys Tyr Glu Glu Ile Cys Ala
                325                 330                 335

cca caa gta gaa gaa ttc tgc tac ata aca gac aat aca tat ttc aga     1056
Pro Gln Val Glu Glu Phe Cys Tyr Ile Thr Asp Asn Thr Tyr Phe Arg
            340                 345                 350

gat gag gtt cta gat atg gaa gct tct gta ctg aat tac ttg aag ttt     1104
Asp Glu Val Leu Asp Met Glu Ala Ser Val Leu Asn Tyr Leu Lys Phe
        355                 360                 365

gaa atg aca gca cct aca gca aag tgc ttt ttg agg aga ttt gct cgt     1152
Glu Met Thr Ala Pro Thr Ala Lys Cys Phe Leu Arg Arg Phe Ala Arg
    370                 375                 380

gct gca caa gcg tgt gat gag gat cct gct ttg cat ctg gag ttc ctt     1200
Ala Ala Gln Ala Cys Asp Glu Asp Pro Ala Leu His Leu Glu Phe Leu
385                 390                 395                 400

gcc aat tac atc gct gag cta tca cta ctg gag tac agt cta ctc tct     1248
Ala Asn Tyr Ile Ala Glu Leu Ser Leu Leu Glu Tyr Ser Leu Leu Ser
                405                 410                 415

tac cct cca tca cta ata gct gca tcg gct att ttc ttg gcg aga ttt     1296
Tyr Pro Pro Ser Leu Ile Ala Ala Ser Ala Ile Phe Leu Ala Arg Phe
            420                 425                 430

ata cta cag cca aca aag tat cct tgg aat tcc aca ctc gct cat tac     1344
Ile Leu Gln Pro Thr Lys Tyr Pro Trp Asn Ser Thr Leu Ala His Tyr
        435                 440                 445

aca cag tac aaa ccg tcc aaa ctg agc gaa tgt gta aag gca ttg cat     1392
Thr Gln Tyr Lys Pro Ser Lys Leu Ser Glu Cys Val Lys Ala Leu His
    450                 455                 460

cgc ctt tgc agc gtt ggt tct ggt agc aac ctt cca gca atc aga gaa     1440
Arg Leu Cys Ser Val Gly Ser Gly Ser Asn Leu Pro Ala Ile Arg Glu
465                 470                 475                 480

aag tac agt caa cat aag tac aaa ttt gtt gca aag aag cag tgt cca     1488
Lys Tyr Ser Gln His Lys Tyr Lys Phe Val Ala Lys Lys Gln Cys Pro
                485                 490                 495

ccg cag atc cct act gaa ttc ttt cgg gac acg aca tgc tag             1530
Pro Gln Ile Pro Thr Glu Phe Phe Arg Asp Thr Thr Cys
            500                 505
```

<210> 26
<211> 509

<212> PRT
<213> Zea mays

<400> 26

```
Met Ser Thr Ile Ala Ala Ser Arg Arg Ser Ser Ser Ser Ser Ser Ser
 1               5                  10                  15
Ala Thr Ala Lys Arg Pro Ala Ile Ala Asp Gly Pro Gly Gly Pro Lys
            20                  25                  30
Ala Thr Ala Ala Gln Ala Lys Lys Arg Val Ala Leu Gly Asn Ile Thr
        35                  40                  45
Asn Val Ala Ala Arg Gly Gly Arg Ala Ser Val Gly Gly Ser Leu Gly
    50                  55                  60
Asn Val Met Ala Pro Thr Ser Ser Ala Lys Leu Asn Pro Thr Val Pro
65                  70                  75                  80
Leu Lys Lys Pro Ser Leu Ala Thr Ser Ala Arg Ser Val Ser Ser Ser
                85                  90                  95
Ile Arg Gly Ser Ala Asp Lys Pro Ala Ser Ile Lys Pro Ala Pro Pro
            100                 105                 110
Val Ala Arg His Gly Ser Ala Thr Gln Arg His Asn Asn Val Pro Pro
        115                 120                 125
Pro Lys Val Pro Thr Ile Ala Asp Val Pro Ser Arg Ala Pro Ala Leu
    130                 135                 140
Val Ser Cys Thr Gly Leu Val Ser Pro Gly Arg Ser Gly Asp Ser Val
145                 150                 155                 160
Ser Ser Asp Glu Thr Met Ser Thr Cys Asp Ser Met Lys Ser Pro Asp
                165                 170                 175
Phe Glu Tyr Val Asp Asn Gln Asp Thr Ser Met Leu Ala Ser Leu Gln
            180                 185                 190
Arg Arg Thr Ser Glu His Leu Arg Ile Ser Glu Asp Arg Asp Val Glu
        195                 200                 205
Glu Asn Lys Arg Lys Lys Asn Ala Val Ala Pro Met Glu Ile Asp Arg
    210                 215                 220
Ile Cys Asp Val Asp Ser Glu Tyr Glu Asp Pro Gln Leu Cys Ala Thr
225                 230                 235                 240
Leu Ala Ser Asp Ile Tyr Met His Leu Arg Glu Ala Glu Thr Lys Lys
                245                 250                 255
Arg Pro Ser Thr Asp Phe Met Glu Thr Ile Gln Lys Asp Val Asn Pro
            260                 265                 270
Ser Met Arg Ala Ile Leu Ile Asp Trp Leu Val Glu Val Ala Glu Glu
        275                 280                 285
Tyr Arg Leu Val Pro Asp Thr Leu Tyr Leu Thr Val Asn Tyr Ile Asp
    290                 295                 300
Arg Tyr Leu Ser Gly Asn Glu Ile Ser Arg Gln Arg Leu Gln Leu Leu
305                 310                 315                 320
Gly Val Ala Cys Met Leu Ile Ala Ala Lys Tyr Glu Glu Ile Cys Ala
                325                 330                 335
Pro Gln Val Glu Glu Phe Cys Tyr Ile Thr Asp Asn Thr Tyr Phe Arg
            340                 345                 350
Asp Glu Val Leu Asp Met Glu Ala Ser Val Leu Asn Tyr Leu Lys Phe
        355                 360                 365
Glu Met Thr Ala Pro Thr Ala Lys Cys Phe Leu Arg Arg Phe Ala Arg
    370                 375                 380
Ala Ala Gln Ala Cys Asp Glu Asp Pro Ala Leu His Leu Glu Phe Leu
385                 390                 395                 400
Ala Asn Tyr Ile Ala Glu Leu Ser Leu Leu Glu Tyr Ser Leu Leu Ser
                405                 410                 415
Tyr Pro Pro Ser Leu Ile Ala Ala Ser Ala Ile Phe Leu Ala Arg Phe
            420                 425                 430
Ile Leu Gln Pro Thr Lys Tyr Pro Trp Asn Ser Thr Leu Ala His Tyr
        435                 440                 445
Thr Gln Tyr Lys Pro Ser Lys Leu Ser Glu Cys Val Lys Ala Leu His
    450                 455                 460
```

```
Arg Leu Cys Ser Val Gly Ser Gly Ser Asn Leu Pro Ala Ile Arg Glu
465                 470                 475                 480
Lys Tyr Ser Gln His Lys Tyr Lys Phe Val Ala Lys Lys Gln Cys Pro
                485                 490                 495
Pro Gln Ile Pro Thr Glu Phe Phe Arg Asp Thr Thr Cys
            500                 505
```

<210> 27
<211> 20
<212> DNA
<213> Zea mays

<400> 27
atgtcgacca tcgctgcctc 20

<210> 28
<211> 20
<212> DNA
<213> Zea mays

<400> 28
ctagcatgtc gtgtcccgaa 20

<210> 29
<211> 1275
<212> DNA
<213> Zea mays

<220>
<221> CDS
<222> (1)...(1272)

<400> 29

```
atg gcg gcg cgg gcg gct gac gag aac agg aga ccg gcg gca ggg aag      48
Met Ala Ala Arg Ala Ala Asp Glu Asn Arg Arg Pro Ala Ala Gly Lys
 1               5                  10                  15

ccc gcg cca ggc gtc cga gac atg gcg agc cgg cgc gcg ctc acg gac      96
Pro Ala Pro Gly Val Arg Asp Met Ala Ser Arg Arg Ala Leu Thr Asp
            20                  25                  30

atc aag aac ctc gtc ggg gct gcc ccg tac ccc tac gcc gtc gcc aag     144
Ile Lys Asn Leu Val Gly Ala Ala Pro Tyr Pro Tyr Ala Val Ala Lys
        35                  40                  45

aag ccc atg ctg cag aag agc aaa agg gac gaa aag cag cca gcg ttg     192
Lys Pro Met Leu Gln Lys Ser Lys Arg Asp Glu Lys Gln Pro Ala Leu
    50                  55                  60

gca agc agc cgg ccc atg aca agg aaa ttc gcc gcc tcc ttg gcg agc     240
Ala Ser Ser Arg Pro Met Thr Arg Lys Phe Ala Ala Ser Leu Ala Ser
65                  70                  75                  80

aag ggc caa cct gaa tgt cag ccg atc gta gct gat cca gaa ccc gaa     288
Lys Gly Gln Pro Glu Cys Gln Pro Ile Val Ala Asp Pro Glu Pro Glu
                85                  90                  95

gtt tgt caa cag aag gaa tca gta ggc gat ggc acc gtt gat att gac     336
Val Cys Gln Gln Lys Glu Ser Val Gly Asp Gly Thr Val Asp Ile Asp
            100                 105                 110

gtg gaa ctc tac gag ctg gtc gac ggt agt gat agt gac atc gac atg     384
```

```
Val Glu Leu Tyr Glu Leu Val Asp Gly Ser Asp Ser Asp Ile Asp Met
        115                 120                 125

ggt gcg aca gag aac aag gac att atg aac gaa gat gaa ttg ctc atg      432
Gly Ala Thr Glu Asn Lys Asp Ile Met Asn Glu Asp Glu Leu Leu Met
    130                 135                 140

gat att gac agt gca gac tcg ggg aac ccg ctt gct gca aca gaa tat      480
Asp Ile Asp Ser Ala Asp Ser Gly Asn Pro Leu Ala Ala Thr Glu Tyr
145                 150                 155                 160

gtt aaa gag ctt tac acc ttt tac aga gaa aat gag gct aag agt tgt      528
Val Lys Glu Leu Tyr Thr Phe Tyr Arg Glu Asn Glu Ala Lys Ser Cys
                165                 170                 175

gta agg cca gat tac atg tcc agc caa caa gac ata aac tca aag atg      576
Val Arg Pro Asp Tyr Met Ser Ser Gln Gln Asp Ile Asn Ser Lys Met
            180                 185                 190

aga gca att ctg att gac tgg ctg att gag gtt cac tac aag ttt gaa      624
Arg Ala Ile Leu Ile Asp Trp Leu Ile Glu Val His Tyr Lys Phe Glu
        195                 200                 205

ctg atg gat gag acg ctc ttt ctt atg gta aac ata ata gat aga ttc      672
Leu Met Asp Glu Thr Leu Phe Leu Met Val Asn Ile Ile Asp Arg Phe
    210                 215                 220

ttg gaa aag gaa gtg gtt cca agg aag aag cta caa ctg gtt gga gtc      720
Leu Glu Lys Glu Val Val Pro Arg Lys Lys Leu Gln Leu Val Gly Val
225                 230                 235                 240

aca gct atg ctg ctc gct tgt aaa tat gag gag gta tct gtt cca gtt      768
Thr Ala Met Leu Leu Ala Cys Lys Tyr Glu Glu Val Ser Val Pro Val
                245                 250                 255

gtt gag gac ctt gtg ctg ata tct gac cgt gcc tac aca aaa ggg caa      816
Val Glu Asp Leu Val Leu Ile Ser Asp Arg Ala Tyr Thr Lys Gly Gln
            260                 265                 270

att tta gaa atg gaa aag ttg att ctg aac acg ctg cag ttc aac atg      864
Ile Leu Glu Met Glu Lys Leu Ile Leu Asn Thr Leu Gln Phe Asn Met
        275                 280                 285

tct gtt cca aca cct tat gtc ttc atg aag agg ttt ctg aaa gct gca      912
Ser Val Pro Thr Pro Tyr Val Phe Met Lys Arg Phe Leu Lys Ala Ala
    290                 295                 300

gat gca gat aaa cag ctt gag cta gcg tca ttt ttc atg ctg gag ctc      960
Asp Ala Asp Lys Gln Leu Glu Leu Ala Ser Phe Phe Met Leu Glu Leu
305                 310                 315                 320

tgc ttg gta gaa tac caa atg ctg aat tat cgg cct tcg cat ctg gct     1008
Cys Leu Val Glu Tyr Gln Met Leu Asn Tyr Arg Pro Ser His Leu Ala
                325                 330                 335

gct gct gcg gtt tat act gca cag tgt gct atc aat cgt tgc cag cac     1056
Ala Ala Ala Val Tyr Thr Ala Gln Cys Ala Ile Asn Arg Cys Gln His
            340                 345                 350

tgg aca aag gtc tgc gag tct cat agc aga tac act agc gac caa ctc     1104
Trp Thr Lys Val Cys Glu Ser His Ser Arg Tyr Thr Ser Asp Gln Leu
        355                 360                 365
```

```
ctg gag tgc tcg agg atg atg gta gat ttt cac cag aag gct gga acc      1152
Leu Glu Cys Ser Arg Met Met Val Asp Phe His Gln Lys Ala Gly Thr
    370                 375                 380

agt aag ctc act ggc gtg cac agg aag tac agt acc tac aag ttc ggt      1200
Ser Lys Leu Thr Gly Val His Arg Lys Tyr Ser Thr Tyr Lys Phe Gly
385                 390                 395                 400

tgc gtg gcc aag att ttg cct gcg cag ttc ctg ctg gag tcg gga ggg      1248
Cys Val Ala Lys Ile Leu Pro Ala Gln Phe Leu Leu Glu Ser Gly Gly
                405                 410                 415

aca ccg cct cct tca ggt gca aac tag                                  1275
Thr Pro Pro Pro Ser Gly Ala Asn
            420
```

<210> 30
<211> 424
<212> PRT
<213> Zea mays

<400> 30

```
Met Ala Ala Arg Ala Ala Asp Glu Asn Arg Arg Pro Ala Ala Gly Lys
 1               5                  10                  15
Pro Ala Pro Gly Val Arg Asp Met Ala Ser Arg Arg Ala Leu Thr Asp
            20                  25                  30
Ile Lys Asn Leu Val Gly Ala Ala Pro Tyr Pro Tyr Ala Val Ala Lys
        35                  40                  45
Lys Pro Met Leu Gln Lys Ser Lys Arg Asp Glu Lys Gln Pro Ala Leu
    50                  55                  60
Ala Ser Ser Arg Pro Met Thr Arg Lys Phe Ala Ala Ser Leu Ala Ser
65                  70                  75                  80
Lys Gly Gln Pro Glu Cys Gln Pro Ile Val Ala Asp Pro Glu Pro Glu
                85                  90                  95
Val Cys Gln Gln Lys Glu Ser Val Gly Asp Gly Thr Val Asp Ile Asp
            100                 105                 110
Val Glu Leu Tyr Glu Leu Val Asp Gly Ser Asp Ser Asp Ile Asp Met
        115                 120                 125
Gly Ala Thr Glu Asn Lys Asp Ile Met Asn Glu Asp Glu Leu Leu Met
    130                 135                 140
Asp Ile Asp Ser Ala Asp Ser Gly Asn Pro Leu Ala Ala Thr Glu Tyr
145                 150                 155                 160
Val Lys Glu Leu Tyr Thr Phe Tyr Arg Glu Asn Glu Ala Lys Ser Cys
                165                 170                 175
Val Arg Pro Asp Tyr Met Ser Ser Gln Gln Asp Ile Asn Ser Lys Met
            180                 185                 190
Arg Ala Ile Leu Ile Asp Trp Leu Ile Glu Val His Tyr Lys Phe Glu
        195                 200                 205
Leu Met Asp Glu Thr Leu Phe Leu Met Val Asn Ile Ile Asp Arg Phe
    210                 215                 220
Leu Glu Lys Glu Val Val Pro Arg Lys Lys Leu Gln Leu Val Gly Val
225                 230                 235                 240
Thr Ala Met Leu Leu Ala Cys Lys Tyr Glu Glu Val Ser Val Pro Val
                245                 250                 255
Val Glu Asp Leu Val Leu Ile Ser Asp Arg Ala Tyr Thr Lys Gly Gln
            260                 265                 270
Ile Leu Glu Met Glu Lys Leu Ile Leu Asn Thr Leu Gln Phe Asn Met
        275                 280                 285
Ser Val Pro Thr Pro Tyr Val Phe Met Lys Arg Phe Leu Lys Ala Ala
    290                 295                 300
```

```
Asp Ala Asp Lys Gln Leu Glu Leu Ala Ser Phe Phe Met Leu Glu Leu
305                 310                 315                 320
Cys Leu Val Glu Tyr Gln Met Leu Asn Tyr Arg Pro Ser His Leu Ala
                325                 330                 335
Ala Ala Ala Val Tyr Thr Ala Gln Cys Ala Ile Asn Arg Cys Gln His
            340                 345                 350
Trp Thr Lys Val Cys Glu Ser His Ser Arg Tyr Thr Ser Asp Gln Leu
        355                 360                 365
Leu Glu Cys Ser Arg Met Met Val Asp Phe His Gln Lys Ala Gly Thr
    370                 375                 380
Ser Lys Leu Thr Gly Val His Arg Lys Tyr Ser Thr Tyr Lys Phe Gly
385                 390                 395                 400
Cys Val Ala Lys Ile Leu Pro Ala Gln Phe Leu Leu Glu Ser Gly Gly
                405                 410                 415
Thr Pro Pro Pro Ser Gly Ala Asn
            420
```

<210> 31
<211> 20
<212> DNA
<213> Zea mays

<400> 31
atggcggcgc gggcggctga 20

<210> 32
<211> 20
<212> DNA
<213> Zea mays

<400> 32
ctagtttgca cctgaaggag 20

<210> 33
<211> 774
<212> DNA
<213> Zea mays

<220>
<221> CDS
<222> (1)...(771)

<400> 33

```
atg gcc gcc aac ttc tgg acg tcg tcg cac tgc aag cag ctg ctg gac      48
Met Ala Ala Asn Phe Trp Thr Ser Ser His Cys Lys Gln Leu Leu Asp
 1               5                  10                  15

ccc gag gac gtg gac ctg gtg ccg gcg gcg gac cgg gag cga ggc atc      96
Pro Glu Asp Val Asp Leu Val Pro Ala Ala Asp Arg Glu Arg Gly Ile
            20                  25                  30

acg ccg gaa gag ttc cgc ctc atc aag atc cat atg tcg ttc cat atc     144
Thr Pro Glu Glu Phe Arg Leu Ile Lys Ile His Met Ser Phe His Ile
        35                  40                  45

tgg agg cta gca cag cag gtt aaa gtt agg caa aga gtt gta gct aca     192
Trp Arg Leu Ala Gln Gln Val Lys Val Arg Gln Arg Val Val Ala Thr
    50                  55                  60

gca ata gct tac ttc agg cgt gtg tat aca agg aag agc atg tca gac     240
Ala Ile Ala Tyr Phe Arg Arg Val Tyr Thr Arg Lys Ser Met Ser Asp
65                  70                  75                  80
```

```
tat gat cct cgt ttg gtt gca cct act tgt ttg tat ttg gca tct aag      288
Tyr Asp Pro Arg Leu Val Ala Pro Thr Cys Leu Tyr Leu Ala Ser Lys
                 85                  90                  95

gtg gag gag agt aca gtg caa gcc aga ctt ctt gta ttt tat ata aaa      336
Val Glu Glu Ser Thr Val Gln Ala Arg Leu Leu Val Phe Tyr Ile Lys
            100                 105                 110

aag atg tgt ggt tcc gat gat aag tat cgg ttt gaa att aag gat att      384
Lys Met Cys Gly Ser Asp Asp Lys Tyr Arg Phe Glu Ile Lys Asp Ile
        115                 120                 125

ctt gaa atg gaa atg aag ctc ctg gaa gct ctc gac tat tat ttg gtt      432
Leu Glu Met Glu Met Lys Leu Leu Glu Ala Leu Asp Tyr Tyr Leu Val
    130                 135                 140

gtt ttc cat cca tat cgt cct ctc tta cag ttg ttg cag gat gct ggc      480
Val Phe His Pro Tyr Arg Pro Leu Leu Gln Leu Leu Gln Asp Ala Gly
145                 150                 155                 160

ata aca gac ctg aca caa ttt gcc tgg ggc ctt gtt aat gat aca tac      528
Ile Thr Asp Leu Thr Gln Phe Ala Trp Gly Leu Val Asn Asp Thr Tyr
                165                 170                 175

aag atg gac ctt atc ctc ata tac cct ccc tac atg att gca ctg gcc      576
Lys Met Asp Leu Ile Leu Ile Tyr Pro Pro Tyr Met Ile Ala Leu Ala
            180                 185                 190

tgc ata tac ata gcg agt gtt ttg aaa gat aag gac acc aca gca tgg      624
Cys Ile Tyr Ile Ala Ser Val Leu Lys Asp Lys Asp Thr Thr Ala Trp
        195                 200                 205

ttt gaa gag ctc cgt gtt gac atg aac ata gtt aag aat atc tct atg      672
Phe Glu Glu Leu Arg Val Asp Met Asn Ile Val Lys Asn Ile Ser Met
    210                 215                 220

gaa ata tta gac ttc tat gac acc tac aag att gat cct caa agg ggc      720
Glu Ile Leu Asp Phe Tyr Asp Thr Tyr Lys Ile Asp Pro Gln Arg Gly
225                 230                 235                 240

atc cct gag gat aag ata agc ccc gtg atg aac aag ttg ccg gca aag      768
Ile Pro Glu Asp Lys Ile Ser Pro Val Met Asn Lys Leu Pro Ala Lys
                245                 250                 255

gct taa                                                             774
Ala
```

<210> 34
<211> 257
<212> PRT
<213> Zea mays

<400> 34

```
Met Ala Ala Asn Phe Trp Thr Ser Ser His Cys Lys Gln Leu Leu Asp
 1               5                  10                  15
Pro Glu Asp Val Asp Leu Val Pro Ala Ala Asp Arg Glu Arg Gly Ile
            20                  25                  30
Thr Pro Glu Glu Phe Arg Leu Ile Lys Ile His Met Ser Phe His Ile
        35                  40                  45
```

```
Trp Arg Leu Ala Gln Gln Val Lys Val Arg Gln Arg Val Val Ala Thr
    50                  55                  60
Ala Ile Ala Tyr Phe Arg Arg Val Tyr Thr Arg Lys Ser Met Ser Asp
65                  70                  75                  80
Tyr Asp Pro Arg Leu Val Ala Pro Thr Cys Leu Tyr Leu Ala Ser Lys
                85                  90                  95
Val Glu Glu Ser Thr Val Gln Ala Arg Leu Leu Val Phe Tyr Ile Lys
            100                 105                 110
Lys Met Cys Gly Ser Asp Asp Lys Tyr Arg Phe Glu Ile Lys Asp Ile
        115                 120                 125
Leu Glu Met Glu Met Lys Leu Leu Glu Ala Leu Asp Tyr Tyr Leu Val
    130                 135                 140
Val Phe His Pro Tyr Arg Pro Leu Leu Gln Leu Leu Gln Asp Ala Gly
145                 150                 155                 160
Ile Thr Asp Leu Thr Gln Phe Ala Trp Gly Leu Val Asn Asp Thr Tyr
                165                 170                 175
Lys Met Asp Leu Ile Leu Ile Tyr Pro Pro Tyr Met Ile Ala Leu Ala
            180                 185                 190
Cys Ile Tyr Ile Ala Ser Val Leu Lys Asp Lys Asp Thr Thr Ala Trp
        195                 200                 205
Phe Glu Glu Leu Arg Val Asp Met Asn Ile Val Lys Asn Ile Ser Met
    210                 215                 220
Glu Ile Leu Asp Phe Tyr Asp Thr Tyr Lys Ile Asp Pro Gln Arg Gly
225                 230                 235                 240
Ile Pro Glu Asp Lys Ile Ser Pro Val Met Asn Lys Leu Pro Ala Lys
                245                 250                 255
Ala
```

<210> 35
<211> 20
<212> DNA
<213> Zea mays

<400> 35
atggccgcca acttctggac 20

<210> 36
<211> 20
<212> DNA
<213> Zea mays

<400> 36
ttaagccttt gccggcaact 20

<210> 37
<211> 1173
<212> DNA
<213> Zea mays

<220>
<221> CDS
<222> (1)...(1170)

<400> 37

```
atg ata tac acg gcg atc gac acg ttc tac ctg aca gat gag cag ctc      48
Met Ile Tyr Thr Ala Ile Asp Thr Phe Tyr Leu Thr Asp Glu Gln Leu
 1               5                  10                  15

cgt gac tcg cca tca cgg aag gat ggg atc gac gag gcc acc gag act      96
Arg Asp Ser Pro Ser Arg Lys Asp Gly Ile Asp Glu Ala Thr Glu Thr
            20                  25                  30
```

```
gcg ctc cgc gtc tat ggc tgt gac ctc atc cag gag agc ggc atc ctc      144
Ala Leu Arg Val Tyr Gly Cys Asp Leu Ile Gln Glu Ser Gly Ile Leu
          35                  40                  45

ctc agg cta cca caa gca gtg atg gcc acg gca caa gta ttg ttc cat      192
Leu Arg Leu Pro Gln Ala Val Met Ala Thr Ala Gln Val Leu Phe His
     50                  55                  60

cgc ttt tac tgc aag aaa tca ttt gtt cga ttt agt gca aag aga gtt      240
Arg Phe Tyr Cys Lys Lys Ser Phe Val Arg Phe Ser Ala Lys Arg Val
 65                  70                  75                  80

gct gct agc tgt gtt tgg ctg gca ggg aag ttg gag gag agt ccc agg      288
Ala Ala Ser Cys Val Trp Leu Ala Gly Lys Leu Glu Glu Ser Pro Arg
                 85                  90                  95

aaa tca agg cat att ata ttt gtc ttc cac aga atg gaa tgt agg aga      336
Lys Ser Arg His Ile Ile Phe Val Phe His Arg Met Glu Cys Arg Arg
            100                 105                 110

gaa aac ttg cca att gaa ttt tta gat gtt ttt tca aag aaa tat tcg      384
Glu Asn Leu Pro Ile Glu Phe Leu Asp Val Phe Ser Lys Lys Tyr Ser
        115                 120                 125

gaa ctg agg cat gac ctg ata cgg aca gaa cgg cat ctg ttg aag gag      432
Glu Leu Arg His Asp Leu Ile Arg Thr Glu Arg His Leu Leu Lys Glu
    130                 135                 140

atg ggg ttt att tgc cat gtc gag cac ccc cat aag ttc ata tca aac      480
Met Gly Phe Ile Cys His Val Glu His Pro His Lys Phe Ile Ser Asn
145                 150                 155                 160

tac ctt gca aca ctt gaa gcc cct cct gag ctt act caa gag gca tgg      528
Tyr Leu Ala Thr Leu Glu Ala Pro Pro Glu Leu Thr Gln Glu Ala Trp
                165                 170                 175

aac ctt gcc aat gat agc tta agg aca act ctg tgt gtg cga ttt aag      576
Asn Leu Ala Asn Asp Ser Leu Arg Thr Thr Leu Cys Val Arg Phe Lys
            180                 185                 190

agt gaa gta gta gca tgc gga gtt gtg tat gct gca gct agg aga cac      624
Ser Glu Val Val Ala Cys Gly Val Val Tyr Ala Ala Ala Arg Arg His
        195                 200                 205

cgg gtt ccc ctt cct gaa gat cct ccg tgg tgg act gtc ttt gat gct      672
Arg Val Pro Leu Pro Glu Asp Pro Pro Trp Trp Thr Val Phe Asp Ala
    210                 215                 220

gat gaa gca gga att cag gaa gtt tgc agg gtt ctt gct cac ctc tac      720
Asp Glu Ala Gly Ile Gln Glu Val Cys Arg Val Leu Ala His Leu Tyr
225                 230                 235                 240

agc ttg ccc aag gcc caa tac att cca gta tat aaa gac aac gat tcc      768
Ser Leu Pro Lys Ala Gln Tyr Ile Pro Val Tyr Lys Asp Asn Asp Ser
                245                 250                 255

ttt act gtt aag aga atc tct gat cta cag gct tca aag gaa agt cca      816
Phe Thr Val Lys Arg Ile Ser Asp Leu Gln Ala Ser Lys Glu Ser Pro
            260                 265                 270

gca agg gct gtt gct agc gat aag ggt acc cct gta ccc tca agt tct      864
```

```
Ala Arg Ala Val Ala Ser Asp Lys Gly Thr Pro Val Pro Ser Ser Ser
        275                 280                 285

agc cag gag aag gat cca gtg gct aag act ata ctg aac aag gtg aag      912
Ser Gln Glu Lys Asp Pro Val Ala Lys Thr Ile Leu Asn Lys Val Lys
    290                 295                 300

gaa aaa agt gat gac gaa ggt aaa cca ttg ccc gct gaa ttt gat gga      960
Glu Lys Ser Asp Asp Glu Gly Lys Pro Leu Pro Ala Glu Phe Asp Gly
305                 310                 315                 320

aaa gaa aac ccg gtg gca aac tca aaa aat gac aag tct gat tct ggt     1008
Lys Glu Asn Pro Val Ala Asn Ser Lys Asn Asp Lys Ser Asp Ser Gly
                325                 330                 335

gtc gac cgg agt cga gaa aga gag aga tca aga ggg cgg gaa cgt gat     1056
Val Asp Arg Ser Arg Glu Arg Glu Arg Ser Arg Gly Arg Glu Arg Asp
            340                 345                 350

gcc agg ggt agg gat tct gat cgt gat agc agg ggt cgc gag tct gat     1104
Ala Arg Gly Arg Asp Ser Asp Arg Asp Ser Arg Gly Arg Glu Ser Asp
        355                 360                 365

cgt gag agg gac cgg aga cgc tgc tct agg gaa aga agt tca gat gac     1152
Arg Glu Arg Asp Arg Arg Arg Cys Ser Arg Glu Arg Ser Ser Asp Asp
    370                 375                 380

tca ttg gtc att gca att tga                                        1173
Ser Leu Val Ile Ala Ile
385                 390
```

<210> 38
<211> 390
<212> PRT
<213> Zea mays

<400> 38

```
Met Ile Tyr Thr Ala Ile Asp Thr Phe Tyr Leu Thr Asp Glu Gln Leu
 1               5                  10                  15
Arg Asp Ser Pro Ser Arg Lys Asp Gly Ile Asp Glu Ala Thr Glu Thr
            20                  25                  30
Ala Leu Arg Val Tyr Gly Cys Asp Leu Ile Gln Glu Ser Gly Ile Leu
        35                  40                  45
Leu Arg Leu Pro Gln Ala Val Met Ala Thr Ala Gln Val Leu Phe His
    50                  55                  60
Arg Phe Tyr Cys Lys Lys Ser Phe Val Arg Phe Ser Ala Lys Arg Val
65                  70                  75                  80
Ala Ala Ser Cys Val Trp Leu Ala Gly Lys Leu Glu Glu Ser Pro Arg
                85                  90                  95
Lys Ser Arg His Ile Ile Phe Val Phe His Arg Met Glu Cys Arg Arg
            100                 105                 110
Glu Asn Leu Pro Ile Glu Phe Leu Asp Val Phe Ser Lys Lys Tyr Ser
        115                 120                 125
Glu Leu Arg His Asp Leu Ile Arg Thr Glu Arg His Leu Leu Lys Glu
    130                 135                 140
Met Gly Phe Ile Cys His Val Glu His Pro His Lys Phe Ile Ser Asn
145                 150                 155                 160
Tyr Leu Ala Thr Leu Glu Ala Pro Pro Glu Leu Thr Gln Glu Ala Trp
                165                 170                 175
Asn Leu Ala Asn Asp Ser Leu Arg Thr Thr Leu Cys Val Arg Phe Lys
            180                 185                 190
```

```
Ser Glu Val Val Ala Cys Gly Val Val Tyr Ala Ala Ala Arg Arg His
        195                 200                 205
Arg Val Pro Leu Pro Glu Asp Pro Pro Trp Trp Thr Val Phe Asp Ala
    210                 215                 220
Asp Glu Ala Gly Ile Gln Glu Val Cys Arg Val Leu Ala His Leu Tyr
225                 230                 235                 240
Ser Leu Pro Lys Ala Gln Tyr Ile Pro Val Tyr Lys Asp Asn Asp Ser
                245                 250                 255
Phe Thr Val Lys Arg Ile Ser Asp Leu Gln Ala Ser Lys Glu Ser Pro
            260                 265                 270
Ala Arg Ala Val Ala Ser Asp Lys Gly Thr Pro Val Pro Ser Ser Ser
        275                 280                 285
Ser Gln Glu Lys Asp Pro Val Ala Lys Thr Ile Leu Asn Lys Val Lys
    290                 295                 300
Glu Lys Ser Asp Asp Glu Gly Lys Pro Leu Pro Ala Glu Phe Asp Gly
305                 310                 315                 320
Lys Glu Asn Pro Val Ala Asn Ser Lys Asn Asp Lys Ser Asp Ser Gly
                325                 330                 335
Val Asp Arg Ser Arg Glu Arg Glu Arg Ser Arg Gly Arg Glu Arg Asp
            340                 345                 350
Ala Arg Gly Arg Asp Ser Asp Arg Asp Ser Arg Gly Arg Glu Ser Asp
        355                 360                 365
Arg Glu Arg Asp Arg Arg Arg Cys Ser Arg Glu Arg Ser Ser Asp Asp
    370                 375                 380
Ser Leu Val Ile Ala Ile
385                 390
```

<210> 39
<211> 20
<212> DNA
<213> Zea mays

<400> 39
atgatataca cggcgatcga 20

<210> 40
<211> 20
<212> DNA
<213> Zea mays

<400> 40
tcaaattgca atgaccaatg 20

<210> 41
<211> 552
<212> DNA
<213> Zea mays

<220>
<221> CDS
<222> (1)...(549)

<400> 41

```
atg acc cgg gga gag aac cag gag tcg cag acc ggg aac gtc tcc gcc      48
Met Thr Arg Gly Glu Asn Gln Glu Ser Gln Thr Gly Asn Val Ser Ala
 1               5                  10                  15

gca tct gcg gct gcg ccg gga ccg aag ccg gca tcg gcg ggg gcc ggg      96
Ala Ser Ala Ala Ala Pro Gly Pro Lys Pro Ala Ser Ala Gly Ala Gly
            20                  25                  30

aag ggc gcg gag ggc cag tcg gtg gtg cgg cgg ctg cag tcg gag ctg     144
```

```
Lys Gly Ala Glu Gly Gln Ser Val Val Arg Arg Leu Gln Ser Glu Leu
         35                  40                  45

atg gcg ctg atg atg ggc ggc gat ccg ggc gtg tcg gcg ttc ccc gag      192
Met Ala Leu Met Met Gly Gly Asp Pro Gly Val Ser Ala Phe Pro Glu
     50                  55                  60

ggg gac aac atg ctc cac tgg gtg ggc acc atc gcg gga tcc gcc ggg      240
Gly Asp Asn Met Leu His Trp Val Gly Thr Ile Ala Gly Ser Ala Gly
 65                  70                  75                  80

acg gcc tac gag ggc acc tcc tac cgc ctc gcg ctg gcc ttc acc gcc      288
Thr Ala Tyr Glu Gly Thr Ser Tyr Arg Leu Ala Leu Ala Phe Thr Ala
                 85                  90                  95

gag tac ccg tac aag ccg ccc aag gtg cgg ttc gac acc ccc tgc ttc      336
Glu Tyr Pro Tyr Lys Pro Pro Lys Val Arg Phe Asp Thr Pro Cys Phe
            100                 105                 110

cac ccc aac gtc gac gtg cac ggc aac atc tgc ctg gac atc ctc cag      384
His Pro Asn Val Asp Val His Gly Asn Ile Cys Leu Asp Ile Leu Gln
        115                 120                 125

gac aag tgg tcc tcc gcc tac gac gtg cgc acc atc ctc ctc tcc atc      432
Asp Lys Trp Ser Ser Ala Tyr Asp Val Arg Thr Ile Leu Leu Ser Ile
    130                 135                 140

cag agc ctg ctc gga gag ccg aac aac gac tcg ccg ctc aac acg cag      480
Gln Ser Leu Leu Gly Glu Pro Asn Asn Asp Ser Pro Leu Asn Thr Gln
145                 150                 155                 160

gcg gcg gcg ctt tgg gcg aac cag gaa gag ttc tgg aag atg gtg gag      528
Ala Ala Ala Leu Trp Ala Asn Gln Glu Glu Phe Trp Lys Met Val Glu
                165                 170                 175

aag ctc tac aag gcc gcc gcg tag                                     552
Lys Leu Tyr Lys Ala Ala Ala
            180
```

<210> 42
<211> 183
<212> PRT
<213> Zea mays

<400> 42

```
Met Thr Arg Gly Glu Asn Gln Glu Ser Gln Thr Gly Asn Val Ser Ala
 1               5                  10                  15
Ala Ser Ala Ala Ala Pro Gly Pro Lys Pro Ala Ser Ala Gly Ala Gly
            20                  25                  30
Lys Gly Ala Glu Gly Gln Ser Val Val Arg Arg Leu Gln Ser Glu Leu
        35                  40                  45
Met Ala Leu Met Met Gly Gly Asp Pro Gly Val Ser Ala Phe Pro Glu
    50                  55                  60
Gly Asp Asn Met Leu His Trp Val Gly Thr Ile Ala Gly Ser Ala Gly
65                  70                  75                  80
Thr Ala Tyr Glu Gly Thr Ser Tyr Arg Leu Ala Leu Ala Phe Thr Ala
                85                  90                  95
Glu Tyr Pro Tyr Lys Pro Pro Lys Val Arg Phe Asp Thr Pro Cys Phe
            100                 105                 110
His Pro Asn Val Asp Val His Gly Asn Ile Cys Leu Asp Ile Leu Gln
        115                 120                 125
```

```
Asp Lys Trp Ser Ser Ala Tyr Asp Val Arg Thr Ile Leu Leu Ser Ile
    130                 135                 140
Gln Ser Leu Leu Gly Glu Pro Asn Asn Asp Ser Pro Leu Asn Thr Gln
145                 150                 155                 160
Ala Ala Ala Leu Trp Ala Asn Gln Glu Glu Phe Trp Lys Met Val Glu
                165                 170                 175
Lys Leu Tyr Lys Ala Ala Ala
            180
```

<210> 43
<211> 20
<212> DNA
<213> Zea mays

<400> 43
atgacccggg gagagaacca 20

<210> 44
<211> 20
<212> DNA
<213> Zea mays

<400> 44
ctacgcggcg gccttgtaga 20

<210> 45
<211> 2248
<212> DNA
<213> Zea mays

<220>
<221> CDS
<222> (28)...(1905)

<400> 45

```
aaaactagca aggtttgtca gagtcta atg gat ccc aaa gct acc tca aca tcc     54
                               Met Asp Pro Lys Ala Thr Ser Thr Ser
                                1               5

aaa acc gat aac atc gat caa atc act atc att gag gag aaa gtc aac     102
Lys Thr Asp Asn Ile Asp Gln Ile Thr Ile Ile Glu Glu Lys Val Asn
 10                  15                  20                  25

aaa atc ggg act gaa cca aca atc cga aaa tac tcc aag ggc aga atg     150
Lys Ile Gly Thr Glu Pro Thr Ile Arg Lys Tyr Ser Lys Gly Arg Met
                30                  35                  40

cta ggc aaa gga ggc ttt gct aag tgt tat gaa gtg acc aat ctt gag     198
Leu Gly Lys Gly Gly Phe Ala Lys Cys Tyr Glu Val Thr Asn Leu Glu
            45                  50                  55

aat aaa aaa gtt ttg gct ggg aag att atc tgt aag gcc tct ttg aca     246
Asn Lys Lys Val Leu Ala Gly Lys Ile Ile Cys Lys Ala Ser Leu Thr
        60                  65                  70

aag agc aga gcc aaa caa aaa ctt att tct gag ata aaa att cat aaa     294
Lys Ser Arg Ala Lys Gln Lys Leu Ile Ser Glu Ile Lys Ile His Lys
    75                  80                  85

tca ctt cgt cac agc aac att gtt gaa ttt gag cac gtg ttt gaa gac     342
Ser Leu Arg His Ser Asn Ile Val Glu Phe Glu His Val Phe Glu Asp
 90                  95                  100                 105
```

```
caa gaa aat gtc tac atc tta tta gag tta tgt ccc aat caa tcc ctt      390
Gln Glu Asn Val Tyr Ile Leu Leu Glu Leu Cys Pro Asn Gln Ser Leu
                110                 115                 120

cat gac ctt atc aag aga aga aag cga ctt aca gaa ata gaa gta caa      438
His Asp Leu Ile Lys Arg Arg Lys Arg Leu Thr Glu Ile Glu Val Gln
            125                 130                 135

tgt tac acc ctg caa cta ata tgc ggt cta aag tat ctt cac agc cgc      486
Cys Tyr Thr Leu Gln Leu Ile Cys Gly Leu Lys Tyr Leu His Ser Arg
        140                 145                 150

aga gtt atc cat aga gat tta aaa ctt gga aat ctc ctt ctc aat gac      534
Arg Val Ile His Arg Asp Leu Lys Leu Gly Asn Leu Leu Leu Asn Asp
    155                 160                 165

aag atg gag ctc aaa ata tgc gac ttt ggt ctt gca gca aaa ctc gaa      582
Lys Met Glu Leu Lys Ile Cys Asp Phe Gly Leu Ala Ala Lys Leu Glu
170                 175                 180                 185

ttt gac ggt gaa aaa aga aaa acc gtt tgt gga act cca aat tac atc      630
Phe Asp Gly Glu Lys Arg Lys Thr Val Cys Gly Thr Pro Asn Tyr Ile
                190                 195                 200

gct cca gaa gtt ata gaa ggc aag gga ggc cat tca tac gaa gtt gat      678
Ala Pro Glu Val Ile Glu Gly Lys Gly Gly His Ser Tyr Glu Val Asp
            205                 210                 215

acc tgg tca ctc ggc gta att ata tac act tta ctt gtt ggc aga cct      726
Thr Trp Ser Leu Gly Val Ile Ile Tyr Thr Leu Leu Val Gly Arg Pro
        220                 225                 230

cca ttt gaa act tct gat gtc aag caa act tac aag aga atc aag gcc      774
Pro Phe Glu Thr Ser Asp Val Lys Gln Thr Tyr Lys Arg Ile Lys Ala
    235                 240                 245

tgt gaa tac agt ttt cct gac cat gtc tca gtt tct gat aca gct aag      822
Cys Glu Tyr Ser Phe Pro Asp His Val Ser Val Ser Asp Thr Ala Lys
250                 255                 260                 265

aac ctc gtc caa aaa atg ttg act cta gat cct tca aag aga ccc tct      870
Asn Leu Val Gln Lys Met Leu Thr Leu Asp Pro Ser Lys Arg Pro Ser
                270                 275                 280

ctt gac gag att ctt caa cat ccc ttc ctc aaa aat gca aac aac atc      918
Leu Asp Glu Ile Leu Gln His Pro Phe Leu Lys Asn Ala Asn Asn Ile
            285                 290                 295

cca aag ttc ctt cca gct tca aca ctt gct tgt cct cct tca act agc      966
Pro Lys Phe Leu Pro Ala Ser Thr Leu Ala Cys Pro Pro Ser Thr Ser
        300                 305                 310

tac tta aat cag ttt gca agc cct gaa aac tca gta aaa gtt ccc tca     1014
Tyr Leu Asn Gln Phe Ala Ser Pro Glu Asn Ser Val Lys Val Pro Ser
    315                 320                 325

cag cct gca cca aaa tct gcc gag gct aca cca cta gca gcc cag aaa     1062
Gln Pro Ala Pro Lys Ser Ala Glu Ala Thr Pro Leu Ala Ala Gln Lys
330                 335                 340                 345

aat gga aga ttt atc aat acc caa gga agc aat atg ttt ggt tct gag     1110
```

```
Asn Gly Arg Phe Ile Asn Thr Gln Gly Ser Asn Met Phe Gly Ser Glu
                350                 355                 360

aag act ttg gtg act agt cct cac tct gca acc aca caa gcc cat acc      1158
Lys Thr Leu Val Thr Ser Pro His Ser Ala Thr Thr Gln Ala His Thr
            365                 370                 375

aat gag aat gtt gta ttg act tct cag tta gat aga cat caa act caa      1206
Asn Glu Asn Val Val Leu Thr Ser Gln Leu Asp Arg His Gln Thr Gln
        380                 385                 390

gga gaa aaa gga tgg aat ttc act aag aca ggc tct tgg cag tca aat      1254
Gly Glu Lys Gly Trp Asn Phe Thr Lys Thr Gly Ser Trp Gln Ser Asn
    395                 400                 405

ttg aac ggc act caa agt gtc aag gga tct tcc cgt cct caa act gtc      1302
Leu Asn Gly Thr Gln Ser Val Lys Gly Ser Ser Arg Pro Gln Thr Val
410                 415                 420                 425

caa caa aaa gga gac tta aag tca gca caa agt ctt aag gct cct gct      1350
Gln Gln Lys Gly Asp Leu Lys Ser Ala Gln Ser Leu Lys Ala Pro Ala
                430                 435                 440

ctc ctt aat aat tta gga agc aga ctt aga gtt tca ggg tca gca gtt      1398
Leu Leu Asn Asn Leu Gly Ser Arg Leu Arg Val Ser Gly Ser Ala Val
            445                 450                 455

gga tca aat aga gga cag gta tta tct gga aat gaa gtc tgg gtt aag      1446
Gly Ser Asn Arg Gly Gln Val Leu Ser Gly Asn Glu Val Trp Val Lys
        460                 465                 470

aag tgg gtt gat tat tca tcg aaa tac gga atg gga tac aac ctt tca      1494
Lys Trp Val Asp Tyr Ser Ser Lys Tyr Gly Met Gly Tyr Asn Leu Ser
    475                 480                 485

aat ggg aca aca gga gtg ttc ttc aat gat aat acc aag ata gtt ttc      1542
Asn Gly Thr Thr Gly Val Phe Phe Asn Asp Asn Thr Lys Ile Val Phe
490                 495                 500                 505

aat caa aaa aca gat caa gtg act tat atc cag aga ggc aag aat gat      1590
Asn Gln Lys Thr Asp Gln Val Thr Tyr Ile Gln Arg Gly Lys Asn Asp
                510                 515                 520

aga cag gat aca gtg act cat tac tcc ctg aca gag tat ccc aaa gat      1638
Arg Gln Asp Thr Val Thr His Tyr Ser Leu Thr Glu Tyr Pro Lys Asp
            525                 530                 535

ctg cag aaa aag atg aca tta tta caa cac ttt aag aag tat ctc gaa      1686
Leu Gln Lys Lys Met Thr Leu Leu Gln His Phe Lys Lys Tyr Leu Glu
        540                 545                 550

ggc agc gag tat gga ggc tct gag agt atc aat gat gga act gag act      1734
Gly Ser Glu Tyr Gly Gly Ser Glu Ser Ile Asn Asp Gly Thr Glu Thr
    555                 560                 565

caa atc ggc gtc tat gtg aaa aaa tgg gtt aag aca aag aac gca act      1782
Gln Ile Gly Val Tyr Val Lys Lys Trp Val Lys Thr Lys Asn Ala Thr
570                 575                 580                 585

ttg ttt aga ctc agc aat aag aca gtg caa gtc cac ttc act gac aga      1830
Leu Phe Arg Leu Ser Asn Lys Thr Val Gln Val His Phe Thr Asp Arg
                590                 595                 600
```

```
act gag atc atc ctg aac tcc gaa aat aag caa gtg act tat gtt gca      1878
Thr Glu Ile Ile Leu Asn Ser Glu Asn Lys Gln Val Thr Tyr Val Ala
            605                 610                 615

aga aag gag aca gaa cca att tcc cct tgagcacagc gctggaaagc          1925
Arg Lys Glu Thr Glu Pro Ile Ser Pro
        620                 625

actaacactg atatgactaa aagactgaaa tatacaaagg atttgctttc tgcaatgata  1985
aatggacctc ctcagcaagg ccagataccg actttgaact tggcatagtg gaataaactt  2045
tactttcact aagtgaaaaa agaagttggt ttaggttaga tgatatatag actctgaaaa  2105
atgatatgca gttgtaattt tactgtatgg tttttattta gttttaactt tttttaacag  2165
atatgtgaac caattatact ccctatttt tttccttcat ataagtctat gaagaagatt  2225
taaaaaaaaa aaaaaaaaaa aaa                                         2248
```

<210> 46
<211> 626
<212> PRT
<213> Zea mays

<400> 46

```
Met Asp Pro Lys Ala Thr Ser Thr Ser Lys Thr Asp Asn Ile Asp Gln
 1               5                  10                  15
Ile Thr Ile Ile Glu Glu Lys Val Asn Lys Ile Gly Thr Glu Pro Thr
            20                  25                  30
Ile Arg Lys Tyr Ser Lys Gly Arg Met Leu Gly Lys Gly Gly Phe Ala
        35                  40                  45
Lys Cys Tyr Glu Val Thr Asn Leu Glu Asn Lys Lys Val Leu Ala Gly
    50                  55                  60
Lys Ile Ile Cys Lys Ala Ser Leu Thr Lys Ser Arg Ala Lys Gln Lys
65                  70                  75                  80
Leu Ile Ser Glu Ile Lys Ile His Lys Ser Leu Arg His Ser Asn Ile
                85                  90                  95
Val Glu Phe Glu His Val Phe Glu Asp Gln Glu Asn Val Tyr Ile Leu
            100                 105                 110
Leu Glu Leu Cys Pro Asn Gln Ser Leu His Asp Leu Ile Lys Arg Arg
        115                 120                 125
Lys Arg Leu Thr Glu Ile Glu Val Gln Cys Tyr Thr Leu Gln Leu Ile
    130                 135                 140
Cys Gly Leu Lys Tyr Leu His Ser Arg Arg Val Ile His Arg Asp Leu
145                 150                 155                 160
Lys Leu Gly Asn Leu Leu Leu Asn Asp Lys Met Glu Leu Lys Ile Cys
                165                 170                 175
Asp Phe Gly Leu Ala Ala Lys Leu Glu Phe Asp Gly Glu Lys Arg Lys
            180                 185                 190
Thr Val Cys Gly Thr Pro Asn Tyr Ile Ala Pro Glu Val Ile Glu Gly
        195                 200                 205
Lys Gly Gly His Ser Tyr Glu Val Asp Thr Trp Ser Leu Gly Val Ile
    210                 215                 220
Ile Tyr Thr Leu Leu Val Gly Arg Pro Pro Phe Glu Thr Ser Asp Val
225                 230                 235                 240
Lys Gln Thr Tyr Lys Arg Ile Lys Ala Cys Glu Tyr Ser Phe Pro Asp
                245                 250                 255
His Val Ser Val Ser Asp Thr Ala Lys Asn Leu Val Gln Lys Met Leu
            260                 265                 270
Thr Leu Asp Pro Ser Lys Arg Pro Ser Leu Asp Glu Ile Leu Gln His
        275                 280                 285
Pro Phe Leu Lys Asn Ala Asn Asn Ile Pro Lys Phe Leu Pro Ala Ser
    290                 295                 300
Thr Leu Ala Cys Pro Pro Ser Thr Ser Tyr Leu Asn Gln Phe Ala Ser
305                 310                 315                 320
```

```
Pro Glu Asn Ser Val Lys Val Pro Ser Gln Pro Ala Pro Lys Ser Ala
                325                 330                 335
Glu Ala Thr Pro Leu Ala Ala Gln Lys Asn Gly Arg Phe Ile Asn Thr
            340                 345                 350
Gln Gly Ser Asn Met Phe Gly Ser Glu Lys Thr Leu Val Thr Ser Pro
        355                 360                 365
His Ser Ala Thr Thr Gln Ala His Thr Asn Glu Asn Val Val Leu Thr
    370                 375                 380
Ser Gln Leu Asp Arg His Gln Thr Gln Gly Glu Lys Gly Trp Asn Phe
385                 390                 395                 400
Thr Lys Thr Gly Ser Trp Gln Ser Asn Leu Asn Gly Thr Gln Ser Val
                405                 410                 415
Lys Gly Ser Ser Arg Pro Gln Thr Val Gln Gln Lys Gly Asp Leu Lys
            420                 425                 430
Ser Ala Gln Ser Leu Lys Ala Pro Ala Leu Leu Asn Asn Leu Gly Ser
        435                 440                 445
Arg Leu Arg Val Ser Gly Ser Ala Val Gly Ser Asn Arg Gly Gln Val
    450                 455                 460
Leu Ser Gly Asn Glu Val Trp Val Lys Lys Trp Val Asp Tyr Ser Ser
465                 470                 475                 480
Lys Tyr Gly Met Gly Tyr Asn Leu Ser Asn Gly Thr Thr Gly Val Phe
                485                 490                 495
Phe Asn Asp Asn Thr Lys Ile Val Phe Asn Gln Lys Thr Asp Gln Val
            500                 505                 510
Thr Tyr Ile Gln Arg Gly Lys Asn Asp Arg Gln Asp Thr Val Thr His
        515                 520                 525
Tyr Ser Leu Thr Glu Tyr Pro Lys Asp Leu Gln Lys Lys Met Thr Leu
    530                 535                 540
Leu Gln His Phe Lys Lys Tyr Leu Glu Gly Ser Glu Tyr Gly Gly Ser
545                 550                 555                 560
Glu Ser Ile Asn Asp Gly Thr Glu Thr Gln Ile Gly Val Tyr Val Lys
                565                 570                 575
Lys Trp Val Lys Thr Lys Asn Ala Thr Leu Phe Arg Leu Ser Asn Lys
            580                 585                 590
Thr Val Gln Val His Phe Thr Asp Arg Thr Glu Ile Ile Leu Asn Ser
        595                 600                 605
Glu Asn Lys Gln Val Thr Tyr Val Ala Arg Lys Glu Thr Glu Pro Ile
    610                 615                 620
Ser Pro
625
```

<210> 47
<211> 20
<212> DNA
<213> Zea mays

<400> 47
atggatccca aagctacctc 20

<210> 48
<211> 20
<212> DNA
<213> Zea mays

<400> 48
tcaaggggaa attggttctg 20

<210> 49
<211> 1876
<212> DNA
<213> Zea mays

<220>
<221> CDS
<222> (71)...(1615)

<400> 49

```
gtcgacccac gcgtccgcga cctgattcgg aaggagggtt gtagccccac cccctctggt     60
tccggcgccc atg gcg gtg gcg gta cct ggc cag ctc aac cta gac gag         109
           Met Ala Val Ala Val Pro Gly Gln Leu Asn Leu Asp Glu
            1               5                  10

gcc ccg tca tgg ggc tcc cgc agc gtc gac tgc ttc gag aag ctc gag       157
Ala Pro Ser Trp Gly Ser Arg Ser Val Asp Cys Phe Glu Lys Leu Glu
     15                  20                  25

cag atc ggc gag ggc act tac ggg caa gtg tac atg gcg aag gag acg       205
Gln Ile Gly Glu Gly Thr Tyr Gly Gln Val Tyr Met Ala Lys Glu Thr
 30                  35                  40                  45

ggc acc aac gag atc gtc gcg ctc aag aag atc cgc atg gac aac gag       253
Gly Thr Asn Glu Ile Val Ala Leu Lys Lys Ile Arg Met Asp Asn Glu
                 50                  55                  60

cgg gaa ggg ttc ccg atc acc gcc ata cgc gag atc aag atc ctg aag       301
Arg Glu Gly Phe Pro Ile Thr Ala Ile Arg Glu Ile Lys Ile Leu Lys
             65                  70                  75

aag ctt cac cac cag aac gtt atc aag cta aag gag atc gtc acc tcc       349
Lys Leu His His Gln Asn Val Ile Lys Leu Lys Glu Ile Val Thr Ser
         80                  85                  90

cca ggc cca gag cgg gac gag caa ggg agg cca att gac ggc aac aag       397
Pro Gly Pro Glu Arg Asp Glu Gln Gly Arg Pro Ile Asp Gly Asn Lys
     95                  100                 105

tac aag ggg agc att tac atg gtt ttc gag tat atg gat cat gac ctc       445
Tyr Lys Gly Ser Ile Tyr Met Val Phe Glu Tyr Met Asp His Asp Leu
110                 115                 120                 125

act ggg ctt gct gat aga cca gga atg cgg ttc act gtg cca cag att       493
Thr Gly Leu Ala Asp Arg Pro Gly Met Arg Phe Thr Val Pro Gln Ile
                130                 135                 140

aag tgt tac atg aag cag ctg ctt aca ggc cta cac tat tgt cac gtc       541
Lys Cys Tyr Met Lys Gln Leu Leu Thr Gly Leu His Tyr Cys His Val
            145                 150                 155

aat caa gtc ctg cat cga gac atc aaa gga tct aac ctc ttg ata gat       589
Asn Gln Val Leu His Arg Asp Ile Lys Gly Ser Asn Leu Leu Ile Asp
        160                 165                 170

aac gag ggt aac tta aag ctt gct gat ttt ggc ctt gca aga tca ttc       637
Asn Glu Gly Asn Leu Lys Leu Ala Asp Phe Gly Leu Ala Arg Ser Phe
    175                 180                 185

tcc agt gat cac aat gga aac ctt acc aac cgt gtg atc act ttg tgg       685
Ser Ser Asp His Asn Gly Asn Leu Thr Asn Arg Val Ile Thr Leu Trp
190                 195                 200                 205

tat aga cct cca gag ttg cta ctt gga agc aca aag tat agt cca gcc       733
Tyr Arg Pro Pro Glu Leu Leu Leu Gly Ser Thr Lys Tyr Ser Pro Ala
                210                 215                 220
```

```
gtt gac atg tgg tca gtg ggc tgt att ttt gca gaa ctg ctc aat ggg      781
Val Asp Met Trp Ser Val Gly Cys Ile Phe Ala Glu Leu Leu Asn Gly
            225                 230                 235

aag cca ata ttg cct gga aag aat gag cca gag cag ctg act aag atc      829
Lys Pro Ile Leu Pro Gly Lys Asn Glu Pro Glu Gln Leu Thr Lys Ile
        240                 245                 250

ttc gag ctt tgt ggt acc cct gac gac aca atc tgg cct ggt gtc aca      877
Phe Glu Leu Cys Gly Thr Pro Asp Asp Thr Ile Trp Pro Gly Val Thr
    255                 260                 265

aaa atg cca tgg tac aac aac ttc aag cct cac cgg cca tta aag aga      925
Lys Met Pro Trp Tyr Asn Asn Phe Lys Pro His Arg Pro Leu Lys Arg
270                 275                 280                 285

cga gtt aaa gac ttc ttt aaa cat ttt gat cgg cat gca ctg gat ctg      973
Arg Val Lys Asp Phe Phe Lys His Phe Asp Arg His Ala Leu Asp Leu
                290                 295                 300

tta gag aag atg ttg act tta gat cca tca cag agg ata tca gca aaa     1021
Leu Glu Lys Met Leu Thr Leu Asp Pro Ser Gln Arg Ile Ser Ala Lys
            305                 310                 315

gat gca ctt gat gcg gaa tat ttc tgg act gat cct tta cct tgc gat     1069
Asp Ala Leu Asp Ala Glu Tyr Phe Trp Thr Asp Pro Leu Pro Cys Asp
        320                 325                 330

cca aaa agc ttg ccc aag tat gag gca tca cat gaa ttc cag act aag     1117
Pro Lys Ser Leu Pro Lys Tyr Glu Ala Ser His Glu Phe Gln Thr Lys
    335                 340                 345

aaa aag cgt cag caa cag agg cag gca gag gaa gct gca aag cgt caa     1165
Lys Lys Arg Gln Gln Gln Arg Gln Ala Glu Glu Ala Ala Lys Arg Gln
350                 355                 360                 365

aaa ctt aat cat cct cca cca cat tct cgc ttg cct cca atc cag cag     1213
Lys Leu Asn His Pro Pro Pro His Ser Arg Leu Pro Pro Ile Gln Gln
                370                 375                 380

cca ggg caa gca cat cct caa atc agg cct ggc cag ggt atg cac aat     1261
Pro Gly Gln Ala His Pro Gln Ile Arg Pro Gly Gln Gly Met His Asn
            385                 390                 395

gtg cct cct gtg gca gct ggg cca ggc cat cac tat aca aag ccc cga     1309
Val Pro Pro Val Ala Ala Gly Pro Gly His His Tyr Thr Lys Pro Arg
        400                 405                 410

gga cca gga ggc cct aac cgg tat cct ccg ggt ggg aat cag ggg gga     1357
Gly Pro Gly Gly Pro Asn Arg Tyr Pro Pro Gly Gly Asn Gln Gly Gly
    415                 420                 425

ggc tac aat ccg aac cgt gga ggt cag ggt ggt ggc tat ggc agt ggc     1405
Gly Tyr Asn Pro Asn Arg Gly Gly Gln Gly Gly Gly Tyr Gly Ser Gly
430                 435                 440                 445

cca tat ccc cag caa ggg cga ggg cct cct tat cct gga ggt gga atg     1453
Pro Tyr Pro Gln Gln Gly Arg Gly Pro Pro Tyr Pro Gly Gly Gly Met
                450                 455                 460

ggt gga aca gct ggt cca cgg ggc agc ggt ggc agt ggc tat gga gct     1501
Gly Gly Thr Ala Gly Pro Arg Gly Ser Gly Gly Ser Gly Tyr Gly Ala
```

```
            465                 470                 475

gga ggt cca aac tac caa cag ggt ggt cca tat ggc gcg tcc ggt cca      1549
Gly Gly Pro Asn Tyr Gln Gln Gly Gly Pro Tyr Gly Ala Ser Gly Pro
        480                 485                 490

ggc cga gga tca aac tac tcc cag ggt ggt ggt tcc cgc aat cag cag      1597
Gly Arg Gly Ser Asn Tyr Ser Gln Gly Gly Gly Ser Arg Asn Gln Gln
    495                 500                 505

cag tat ggg aac tgg caa taacgtggca ttgagatgta tgtatatgca             1645
Gln Tyr Gly Asn Trp Gln
510                 515

gaatgcattt gcagcatctt tgcagtttaa tcatgttgtg attagtactg taatacagtc    1705
cagatatcgt aatagggctg tttgtggcaa gcccgagcgt tgccaaaggg acccttagca    1765
tagtctcatt tttacctcat cacaaaacca acttaaacac aataatagta aactttccgt    1825
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa agggcggccg c             1876

      <210> 50
      <211> 515
      <212> PRT
      <213> Zea mays

      <400> 50
Met Ala Val Ala Val Pro Gly Gln Leu Asn Leu Asp Glu Ala Pro Ser
 1               5                  10                  15
Trp Gly Ser Arg Ser Val Asp Cys Phe Glu Lys Leu Glu Gln Ile Gly
            20                  25                  30
Glu Gly Thr Tyr Gly Gln Val Tyr Met Ala Lys Glu Thr Gly Thr Asn
        35                  40                  45
Glu Ile Val Ala Leu Lys Lys Ile Arg Met Asp Asn Glu Arg Glu Gly
    50                  55                  60
Phe Pro Ile Thr Ala Ile Arg Glu Ile Lys Ile Leu Lys Lys Leu His
65                  70                  75                  80
His Gln Asn Val Ile Lys Leu Lys Glu Ile Val Thr Ser Pro Gly Pro
                85                  90                  95
Glu Arg Asp Glu Gln Gly Arg Pro Ile Asp Gly Asn Lys Tyr Lys Gly
            100                 105                 110
Ser Ile Tyr Met Val Phe Glu Tyr Met Asp His Asp Leu Thr Gly Leu
        115                 120                 125
Ala Asp Arg Pro Gly Met Arg Phe Thr Val Pro Gln Ile Lys Cys Tyr
    130                 135                 140
Met Lys Gln Leu Leu Thr Gly Leu His Tyr Cys His Val Asn Gln Val
145                 150                 155                 160
Leu His Arg Asp Ile Lys Gly Ser Asn Leu Leu Ile Asp Asn Glu Gly
                165                 170                 175
Asn Leu Lys Leu Ala Asp Phe Gly Leu Ala Arg Ser Phe Ser Ser Asp
            180                 185                 190
His Asn Gly Asn Leu Thr Asn Arg Val Ile Thr Leu Trp Tyr Arg Pro
        195                 200                 205
Pro Glu Leu Leu Leu Gly Ser Thr Lys Tyr Ser Pro Ala Val Asp Met
    210                 215                 220
Trp Ser Val Gly Cys Ile Phe Ala Glu Leu Leu Asn Gly Lys Pro Ile
225                 230                 235                 240
Leu Pro Gly Lys Asn Glu Pro Glu Gln Leu Thr Lys Ile Phe Glu Leu
                245                 250                 255
Cys Gly Thr Pro Asp Asp Thr Ile Trp Pro Gly Val Thr Lys Met Pro
            260                 265                 270
Trp Tyr Asn Asn Phe Lys Pro His Arg Pro Leu Lys Arg Arg Val Lys
        275                 280                 285
Asp Phe Phe Lys His Phe Asp Arg His Ala Leu Asp Leu Leu Glu Lys
```

```
    290                 295                 300
Met Leu Thr Leu Asp Pro Ser Gln Arg Ile Ser Ala Lys Asp Ala Leu
305                 310                 315                 320
Asp Ala Glu Tyr Phe Trp Thr Asp Pro Leu Pro Cys Asp Pro Lys Ser
                325                 330                 335
Leu Pro Lys Tyr Glu Ala Ser His Glu Phe Gln Thr Lys Lys Lys Arg
            340                 345                 350
Gln Gln Gln Arg Gln Ala Glu Glu Ala Ala Lys Arg Gln Lys Leu Asn
        355                 360                 365
His Pro Pro Pro His Ser Arg Leu Pro Pro Ile Gln Gln Pro Gly Gln
    370                 375                 380
Ala His Pro Gln Ile Arg Pro Gly Gln Gly Met His Asn Val Pro Pro
385                 390                 395                 400
Val Ala Ala Gly Pro Gly His His Tyr Thr Lys Pro Arg Gly Pro Gly
                405                 410                 415
Gly Pro Asn Arg Tyr Pro Pro Gly Gly Asn Gln Gly Gly Gly Tyr Asn
            420                 425                 430
Pro Asn Arg Gly Gly Gln Gly Gly Gly Tyr Gly Ser Gly Pro Tyr Pro
        435                 440                 445
Gln Gln Gly Arg Gly Pro Pro Tyr Pro Gly Gly Gly Met Gly Gly Thr
    450                 455                 460
Ala Gly Pro Arg Gly Ser Gly Gly Ser Gly Tyr Gly Ala Gly Gly Pro
465                 470                 475                 480
Asn Tyr Gln Gln Gly Gly Pro Tyr Gly Ala Ser Gly Pro Gly Arg Gly
                485                 490                 495
Ser Asn Tyr Ser Gln Gly Gly Gly Ser Arg Asn Gln Gln Gln Tyr Gly
            500                 505                 510
Asn Trp Gln
        515
```

<210> 51
<211> 21
<212> DNA
<213> Zea mays

<400> 51
atggcggtgg cggtacctgg c 21

<210> 52
<211> 20
<212> DNA
<213> Zea mays

<400> 52
ttattgccag ttcccatact 20

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
2 November 2000 (02.11.2000)

PCT

(10) International Publication Number
WO 00/65040 A3

(51) International Patent Classification[7]: C07K 14/415, C12N 9/12, 15/54, 5/14

(21) International Application Number: PCT/US00/09975

(22) International Filing Date: 13 April 2000 (13.04.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/130,849 22 April 1999 (22.04.1999) US

(71) Applicant: PIONEER HI-BRED INTERNATIONAL, INC. [US/US]; 800 Capital Square, 400 Locust Street, Des Moines, IA 50309 (US).

(72) Inventors: HELENTJARIS, Timothy, G.; 2960 N.W. 73rd Lane, Ankeny, IA 50021 (US). HABBEN, Jeffrey, E.; 8813 Sunflower Circle, Urbandale, IA 50322 (US). SUN, Yuejin; 5768 N.W. 92nd Street, Johnston, IA 50131 (US).

(74) Agents: VARLEY, Karen, K. et al.; 7100 N.W. 62nd Avenue, Darwin Building, Johnston, IA 50131-1000 (US).

(81) Designated States *(national)*: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *with (an) indication(s) in relation to deposited biological material furnished under Rule 13bis separately from the description*

(88) Date of publication of the international search report:
19 July 2001

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 00/65040 A3

(54) Title: CELL CYCLE GENES FROM PLANTS AND METHODS OF USE

(57) Abstract: The invention provides isolated nucleic acids and their encoded proteins which are involved in cell cycle regulation. The present invention provides methods and compositions relating to altering cyclin and/or cyclin-dependent kinase concentration and/or composition of plants. The invention further provides recombinant expression cassettes, host cells, and transgenic plants.

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C07K14/415 C12N9/12 C12N15/54 C12N5/14

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	DATABASE GENEMBL [Online] 5 November 1997 (1997-11-05) LIN ET AL.: "Arabidopsis thaliana chromosome II section 184 of 255 of the complete sequence." XP002151952 Accession AC003033 ---	1-7,16
A	DATABASE SWISSPROT [Online] 1 June 1998 (1998-06-01) ROUNSLEY ET AL.: "Putative cyclin G-associated kinase from Arabidopsis thaliana" XP002151953 Accession O50071 --- -/--	1-7,16

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
15 November 2000	26. 1. 01

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	ALCONADA RODRIG.., A

4

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 92 09685 A (UNIV AUSTRALIAN) 11 June 1992 (1992-06-11) claims 1-28 ---	8-15
A	WO 98 41642 A (VEYLDER LIEVEN DE ;VLAAMS INTERUNIV INST BIOTECH (BE); INZE DIRK () 24 September 1998 (1998-09-24) page 11, line 20 -page 12, line 2 page 13, line 5-14 page 15, line 5-10 examples 2,7,11-14 claims 16-23 ---	8-15
A	RENAUDIN JEAN-PIERRE ET AL: "Cloning of four cyclins from maize indicates that higher plants have three structurally distinct groups of mitotic cyclins." PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES, vol. 91, no. 15, 1994, pages 7375-7379, XP002151950 1994 ISSN: 0027-8424 page 7375, right-hand column, last paragraph -page 7376, left-hand column, paragraph 3; figure 1; tables 1,2 ---	
A	COLASANTI J ET AL: "ISOLATION AND CHARACTERIZATION OF COMPLEMENTARY DNA CLONES ENCODING A FUNCTIONAL P34-CDC-2 HOMOLOGUE FROM ZEA-MAYS" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES, vol. 88, no. 8, 1991, pages 3377-3381, XP002151951 1991 ISSN: 0027-8424 page 3378, left-hand column, paragraph 3 -page 3379, right-hand column, paragraph 1; figures 1,2 ---	
A	HSIEH WEN-LING ET AL: "Isolation and characterization of a functional A-type cyclin from maize." PLANT MOLECULAR BIOLOGY, vol. 37, no. 1, May 1998 (1998-05), pages 121-129, XP000960369 ISSN: 0167-4412 -----	

4

INTERNATIONAL SEARCH REPORT

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-16 (partially)

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

1. Claims: 1-16 (partially)

An isolated nucleic acid from Zea mays having at least 80% identity with the polynucleotide of SEQ ID NO:1, which codes for a polypeptide of SEQ ID NO:2, a polynucleotide that hybridizes with or that is complementary to the polynucleotide of SEQ ID NO:1 and a polynucleotide comprising at least 25 contiguous nucleotides from SEQ ID NO:1; a recombinant expression cassette containing said polynucleotide and a host cell and a transgenic plant and seed comprising said expression cassette; a method of modulating the level of cell cycle activity in a plant cell by expression of the polynuclelotide of the invention; a polypeptide comprising at least 20 contiguous amino acids from a polypeptide of SEQ ID NO:2, the polypeptide of SEQ ID NO:2 or any polypeptide having at least 80% sequence similarity with the polypeptide of SEQ ID NO:2; a polypeptide encoded by the polynucleotide of SEQ ID NO:1.

2. Claims: 1-16 (partially)

As invention 1, but comprising the polynucleotide of SEQ ID NO:5 and the polypeptide of SEQ ID NO: 6.

3. Claims: 1-16 (partially)

As invention 1, but comprising the polynucleotide of SEQ ID NO:9 and the polypeptide of SEQ ID NO: 10.

4. Claims: 1-16 (partially)

As invention 1, but comprising the polynucleotide of SEQ ID NO:13 and the polypeptide of SEQ ID NO: 14.

5. Claims: 1-16 (partially)

As invention 1, but comprising the polynucleotide of SEQ ID NO:17 and the polypeptide of SEQ ID NO: 18.

6. Claims: 1-16 (partially)

As invention 1, but comprising the polynucleotide of SEQ ID NO:21 and the polypeptide of SEQ ID NO: 22.

7. Claims: 1-16 (partially)

As invention 1, but comprising the polynucleotide of SEQ ID NO:25 and the polypeptide of SEQ ID NO:26.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

8. Claims: 1-16 (partially)

As invention 1, but comprising the polynucleotide of SEQ ID NO:29 and the polypeptide of SEQ ID NO:30.

9. Claims: 1-16 (partially)

As invention 1, but comprising the polynucleotide of SEQ ID NO:33 and the polypeptide of SEQ ID NO:34.

10. Claims: 1-16 (partially)

As invention 1, but comprising the polynucleotide of SEQ ID NO:37 and the polypeptide of SEQ ID NO:38.

11. Claims: 1-16 (partially)

As invention 1, but comprising the polynucleotide of SEQ ID NO:41 and the polypeptide of SEQ ID NO:42.

12. Claims: 1-16 (partially)

As invention 1, but comprising the polynucleotide of SEQ ID NO:45 and the polypeptide of SEQ ID NO:46.

13. Claims: 1-16 (partially)

As invention 1, but comprising the polynucleotide of SEQ ID NO:49 and the polypeptide of SEQ ID NO:50.

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9209685 A	11-06-1992	AU 657722 B	23-03-1995
		AU 9046291 A	25-06-1992
		CA 2097286 A	30-05-1992
		EP 0559729 A	15-09-1993
		JP 6504430 T	26-05-1994
		US 5750862 A	12-05-1998
		US 6087175 A	11-07-2000
WO 9841642 A	24-09-1998	AU 6730198 A	12-10-1998
		EP 0972060 A	19-01-2000